



TITLE:

統計力学と情報処理:自由エネルギーの生み出す新しい情報処理技術
(第48回物性若手夏の学校(2003年度),講義ノート)

AUTHOR(S):

田中, 和之

CITATION:

田中, 和之. 統計力学と情報処理:自由エネルギーの生み出す新しい情報処理技術(第48回物性若手夏の学校(2003年度),講義ノート). 物性研究 2004, 81(5): 679-710

ISSUE DATE:

2004-02-20

URL:

<http://hdl.handle.net/2433/97754>

RIGHT:

統計力学と情報処理

— 自由エネルギーの生み出す新しい情報処理技術 —

東北大学 大学院情報科学研究科 田中 和之¹

概要

ここ 10 年ほどの間に確率的情報処理あるいは情報統計力学といわれる研究テーマが物理学・統計学・情報科学の境界領域を舞台として様々の分野の研究者により研究され、徐々に認知されるようになってきている。何故、統計力学と情報科学が関係するのか？そもそも統計力学は主として物質の性質の解明と予言を目的として発展した学問体系であり、情報科学はデータからの情報の抽出と加工を目的としたものである。しかし、おおざっぱに見て、どちらもたくさんの基本要素が互に関連しあいながら集まって全体を構成している点で、両者の理論的枠組みには多くの類似点がある。この基本的な類似点に着目しながらベイズの公式と呼ばれる学部の確率・統計で教わる簡単な公式を用いることにより、実際の情報処理に役に立つアルゴリズムが提案されつつある。そもそも物理は物質すなわちモノを対象に制限しなければならないという理由はどこにもなく、扱う対象は自由であり、様々の現象に着目し、その本質を見極めることにこそ真の目的がある学問体系である。これまで、たまたまその研究対象としてモノを扱うことが多かっただけであり、コトの物理への拡張は何も今に始まった概念ではない。コトの物理を見極めることにより、情報科学の世界に新しい手法を提供しながら、逆に情報科学における新しい問題を通してモノの物理を鍛えることにより、物理学における新しい見方を開拓してゆくことが期待される。

本講義では確率的情報処理への統計力学のアプローチについての画像修復および人工知能における確率的推論を例にとり、確率的情報処理への統計力学のアプローチの基礎、特に、ベイズ統計とイジング模型を用いた理論的枠組み、自由エネルギー最小の変分原理にもとづく平均場理論を用いた確率的情報処理技術の基礎についてやさしく紹介する。

1 はじめに

物理学の基本的な理念の一つとして自然界に起こる出来事の本質を理解したいということがある。これは物理学にあこがれをいだき大学に入学した方なら一度は多かれ少なかれ考えることであろう。この本質を見極める力こそが、現代の情報処理の時代に置いて重要となりつつある。1990 年代の情報産業はより高速の演算処理とより大容量の記憶媒体の開発にしのぎを削った。そして 21 世紀を迎え IT バブルの崩壊にいたり、量と速さから質に注目が移るようになる。すなわち、個人個人の家庭に携帯電話・コンピュータが半ば家電製品として入り込み、インターネットを通して膨大なデータが個人から個人へとやりとりされる中で、質が問われる時代とはすなわち膨大なデータの中から如何に効率よく必要なデータを引き出し、処理することができるかが問われる時代と考えることができる。このことは毎日の受信数が数十通、数百通にもおよぶ電子メールを読み、処理していて誰もが感ずることであろう。実はこの本質を見極める物理学の考え方、特に統計力学が情報処理に非常に有効であることが最近のいくつかの研究を通して徐々に認知されつつある。

¹E-mail: kazu@statp.is.tohoku.ac.jp

統計力学と情報処理についておおざっぱに言えば、統計力学は主として物質の性質を理解し、予言することを目的として発展してきた学問体系である。これに対して情報処理はデータからの情報の抽出と加工を目的として発達してきた技術体系である。すなわち統計力学の主たる研究対象はモノであり、情報処理の取り扱う対象はコトであるということである。両者は一見全く関連性のないように見えるが、この2つの学問体系の中の一体どこに共通点があるのか？ 実は単純なところに存在する。それは取り扱う対象がどちらも「たくさんの基本要素が関連しながら集まって構成されている」という点にある。物質は原子あるいは分子が相互作用しながらたくさん集まって構成されている。データはビットが決められた並びに従ってたくさん集まることにより意味のある形で構成される。この2つの実に単純な共通点だけをもって、はたして統計力学を情報処理に応用できるのだろうか？ そのように感じる方が多いであろう。しかし、統計力学を情報処理に応用することに成功した問題の多くの本質はこの点に帰着されるのである。統計力学と情報処理の間には統計学という数学を通してみると様々の共通の数理的構造がみえてくる。例えば、統計力学における自由エネルギー、ギブス分布、平均場理論は情報処理においてカルバック・ライブラー情報量、指数関数分布族、確率伝搬法という形で名前を変えて登場する。スピングラス理論の一部はシャノン限界と対応づけられる。これらがいずれも情報処理における基本的概念であることを考えれば統計力学が情報処理技術に有効であることはある程度理解していただけるであろう。

本講義は3日間にわたって行われる。第1日目は、冒頭で統計力学を用いた情報処理の研究についての最近の動向について簡単に紹介し、その後、確率とベイズ統計の基礎について概説し、自由エネルギーと平均場近似の情報論的理解について変分原理の立場でわかりやすく紹介する。第2日目は確率的画像処理について紹介し、ベータ近似と呼ばれる統計力学的手法を通して構成されるアルゴリズムについて、いくつかの数値実験例とともに紹介する。ベータ近似は平均場近似の拡張版であり、本講義では自由エネルギー最小の変分原理から導出する。第3日目は人工知能のひとつの方法論である確率的推論のアルゴリズムを自由エネルギー最小の変分原理にもとづく平均場理論の枠組みから導出し、具体的数値実験の例を紹介する。

2 最近の動向

本節では、統計力学を用いた情報処理の研究の最近の動向について概観する。

統計力学が情報処理に有効であるものとして認知された最初の例はおそらく符号理論であろう。誤り訂正符号から始まり、最近では符号圧縮から暗号理論にいたるまで有効性が指摘されている [1, 2, 3]。その理由は定式化の過程で導出された確率モデルが統計力学におけるある種の物理モデルと数学的構造において多くの類似点を持つ点にある。スピングラス理論を用いた解析的性能評価、平均場理論を用いた復号アルゴリズムの構成など多くの研究成果が報告されている。

画像処理は画素が規則的に配置されていることから、磁性体の物性の解明を目的として開発されてきた統計力学における計算手法の格好の応用の対象である。実際、スピングラス理論、平均場理論をはじめとする多くの統計力学的計算手法がすでに多くの画像処理の問題へと応用されている [4, 5, 7, 6]。また、その一方で量子統計力学的モデルを用いた画像処理の研究への発展もなされている [8]。

人工知能において、ベイズ統計にもとづいて構成された確率的推論システムをベイジアンネットと呼ぶことがある.[9] このベイジアンネットはある種のグラフ表現をもつ確率モデルとして与えられ、グラフ表現が特殊な構造(具体的には木構造)を持つ場合にはピリーフプロパゲーション(信念伝搬法あるいは確率伝搬法)と呼ばれるアルゴリズムを用いて推論を行うことができるということが人工知能の分野で知られていた. その理論的構造は統計力学における転送行列法とほぼ等価であることが次第に明らかになりつつあり、最近では、このピリーフプロパゲーションの枠組みが平均場理論のひとつであるベータ近似およびクラスター変分法を用いて系統的に拡張できることが明らかとなり、ピリーフプロパゲーションをテーマとしたワークショップも開催されるようになってきている [10].

移動体通信における復調方式に統計力学的計算手法を持ち込もうという試みが始まったのは最近2, 3年のことである [11, 12]. 符号分割多元接続(CDMA)方式についてその解析的性能評価にスピングラス理論を導入することに成功したというのが最初の仕事である. 最近, 平均場理論, 確率伝搬法等を用いたアルゴリズムの開発も行われつつある [10].

更に, 機械学習 [13, 14], 遺伝的アルゴリズム [15], インターネットにおけるパケット流のルーティング制御 [16] 等へとベイズ統計・統計力学を応用する研究も行われている. また, 統計力学的計算手法を, 計算論的な立場 [17] からアルゴリズムとして見た研究も行われつつある.

統計力学を情報処理に応用する研究において特に強力な武器となっているのが平均場理論とスピングラス理論である. 平均場理論, スピングラス理論と情報処理についての最近の研究成果を集めた解説書と啓蒙書が文献 [18, 19] という形で出版されている. 更に情報統計物理学という更に広い枠組みで基礎的部分から最近の動向まで詳しく書かれた啓蒙書 [20, 21, 22] も出版されている.

3 確率モデルとベイズ統計

情報処理に統計力学を応用するわけであるから, 当然, 確率モデルを用いた推定へと問題を定式化する必要がある. その際, キーになるのがベイズの公式である. 本節では, ベイズ統計を用いた情報処理の基礎について概説する.

ある事象を考え, その事象として起こりうるすべての場合が合計で M 個として, それに $1, 2, \dots, M$ などの番号を付けたとする. この番号の中のどれかをとり変数 A を導入し, 「例えば 1 番という番号の付けられた事象が起こったことを “ $A = 1$ ” という数学的記号を用いて表すことにした」とする. この時, A を確率変数といい, “ $A = 1$ ” で表現された事象の起こる確率を $\text{Pr}\{A = 1\}$ という記号を用いて表すことにする. 一般に確率変数がその実現値 $1, 2, \dots, M$ のなかの様々な値をとりうるような場合を考えると, $A = a$ ($a = 1, 2, \dots, M$) により指定された事象の確率は $\text{Pr}\{A = a\}$ という表現により与えられる. 確率 $\text{Pr}\{A = a\}$ は

$$\text{Pr}\{A = a\} \geq 0 \quad (a = 1, 2, \dots, M), \quad \sum_{z=1}^M \text{Pr}\{A = z\} = 1 \quad (1)$$

という条件を満たさなければならない. 確率変数 A の確率が

$$\text{Pr}\{A = a\} = P(a) \quad (2)$$

という形で a の関数 $P(a)$ により与えられたとき, この $P(a)$ を確率変数 A の確率分布という. 式 (1) から確率分布は

$$P(a) \geq 0 \quad (a = 1, 2, \dots, M), \quad \sum_{z=1}^M P(z) = 1 \quad (3)$$

を満たさなければならない.

次に, 2 つの事象を考え, その確率変数を A_1, A_2 とし, それぞれの事象として起こりうる場合の総数がそれぞれ M_1 個および M_2 個であるとする. このとき 「 $A_1 = a_1$ 」, 「 $A_2 = a_2$ 」により与えられた事象が両方起きる, すなわち 「 $(A_1 = a_1) \cup (A_2 = a_2)$ 」である確率

$$\Pr\{A_1 = a_1, A_2 = a_2\} \quad (a_1 = 1, 2, \dots, M_1; a_2 = 1, 2, \dots, M_2) \quad (4)$$

を確率変数 A_1 と A_2 に対する結合確率と呼ぶ. 結合確率 $\Pr\{A_1 = a_1, A_2 = a_2\}$ を最初に定義した上で A_2 としてどの事象が起こるかということとは無関係に A_1 が起こる確率を考えた場合, これは

$$\Pr\{A_1 = a_1\} = \sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \delta_{a_1, z_1} \Pr\{A_1 = z_1, A_2 = z_2\} = \sum_{z_2=1}^{M_2} \Pr\{A_1 = a_1, A_2 = z_2\} \quad (5)$$

$$\Pr\{A_2 = a_2\} = \sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \delta_{a_2, z_2} \Pr\{A_1 = z_1, A_2 = z_2\} = \sum_{z_1=1}^{M_1} \Pr\{A_1 = z_1, A_2 = a_2\} \quad (6)$$

と与えられる. $\delta_{a,b} \equiv 1$ ($a = b$), $\delta_{a,b} \equiv 0$ ($a \neq b$) はクロネッカーのデルタである. この時, $\Pr\{A_k = a_k\}$ ($k = 1, 2$) を $\Pr\{A_1 = a_1, A_2 = a_2\}$ の確率変数 A_k についての周辺確率と呼ぶ.

話を一気に一般化してみよう. いま K 個の事象を考え, その確率変数を A_k とし, それぞれの事象として起こりうる場合の総数がそれぞれ M_k 個であるとする. 事象 「 $(A_1 = a_1) \cup (A_1 = a_1) \cup \dots \cup (A_K = a_K)$ 」が起こる結合確率は次のように与えられる.

$$\Pr\{A_1 = a_1, A_2 = a_2, \dots, A_K = a_K\} \quad (a_k = 1, 2, \dots, M_k, k = 1, 2, \dots, K) \quad (7)$$

以後, 確率変数の集合 $\{A_k | k = 1, 2, \dots, K\}$ およびその実現値 $\{a_k | k = 1, 2, \dots, K\}$ をそれぞれ \mathbf{A}, \mathbf{a} という記号で表すことにすると結合確率は $\Pr\{\mathbf{A} = \mathbf{a}\}$ と表される. 結合確率 $\Pr\{\mathbf{A} = \mathbf{a}\}$ は

$$\Pr\{\mathbf{A} = \mathbf{a}\} \geq 0 \quad (a_k = 1, 2, \dots, M_k, k = 1, 2, \dots, K), \quad \sum_{\mathbf{z}} \Pr\{\mathbf{A} = \mathbf{z}\} = 1 \quad (8)$$

という条件を満たさなければならない. ここで, $\sum_{\mathbf{z}}$ は $\mathbf{z} \equiv \{z_k | k = 1, 2, \dots, K\}$ のすべての z_k に対する和を意味する.

$$\sum_{\mathbf{z}} \equiv \sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \dots \sum_{z_K=1}^{M_K} \quad (9)$$

確率変数 \mathbf{A} の確率が

$$\Pr\{\mathbf{A} = \mathbf{a}\} = P(\mathbf{a}) \quad (10)$$

という形で \mathbf{a} の関数 $P(\mathbf{a})$ により与えられたとき, この $P(\mathbf{a})$ を確率変数 \mathbf{A} の結合確率分布という. 式 (8) から結合確率分布 $P(\mathbf{a})$ は

$$P(\mathbf{a}) \geq 0, \quad \sum_{\mathbf{z}} P(\mathbf{z}) = 1 \quad (11)$$

を満たさなければならない. 結合確率 $\Pr\{\mathbf{A} = \mathbf{a}\}$ に対して

$$\Pr\{A_k = a_k\} = \sum_{\mathbf{z}} \delta_{a_k, z_k} \Pr\{\mathbf{A} = \mathbf{z}\} \quad (12)$$

$$\Pr\{A_k = a_k, A_{k'} = a_{k'}\} = \sum_{\mathbf{z}} \delta_{a_k, z_k} \delta_{a_{k'}, z_{k'}} \Pr\{\mathbf{A} = \mathbf{z}\} \quad (13)$$

$$\Pr\{A_k = a_k, A_{k'} = a_{k'}, A_{k''} = a_{k''}\} = \sum_{\mathbf{z}} \delta_{a_k, z_k} \delta_{a_{k'}, z_{k'}} \delta_{a_{k''}, z_{k''}} \Pr\{\mathbf{A} = \mathbf{z}\} \quad (14)$$

という形で様々の周辺確率を定義することができる.

結合確率 $\Pr\{A_1 = a_1, A_2 = a_2\}$ から条件付き確率 $\Pr\{A_1 = a_1 | A_2 = a_2\}$ および $\Pr\{A_2 = a_2 | A_1 = a_1\}$ は次のように定義される.

$$\begin{aligned} \Pr\{A_1 = a_1 | A_2 = a_2\} &\equiv \frac{\Pr\{A_1 = a_1, A_2 = a_2\}}{\Pr\{A_2 = a_2\}} \\ \Pr\{A_2 = a_2 | A_1 = a_1\} &\equiv \frac{\Pr\{A_1 = a_1, A_2 = a_2\}}{\Pr\{A_1 = a_1\}} \end{aligned} \quad (15)$$

この式から

$$\begin{aligned} \Pr\{A_1 = a_1, A_2 = a_2\} &= \Pr\{A_1 = a_1 | A_2 = a_2\} \Pr\{A_2 = a_2\} \\ &= \Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\} \end{aligned} \quad (16)$$

という式が導かれる. 両辺を $\Pr\{A_2 = a_2\}$ で割ることにより次の等式が与えられる.

$$\Pr\{A_1 = a_1 | A_2 = a_2\} = \frac{\Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\}}{\Pr\{A_2 = a_2\}} \quad (17)$$

ここで更に周辺確率の定義から

$$\Pr\{A_2 = a_2\} = \sum_{a_1=1}^{M_1} \Pr\{A_1 = a_1, A_2 = a_2\} = \sum_{a_1=1}^{M_1} \Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\} \quad (18)$$

であることを考慮すると

$$\Pr\{A_1 = a_1 | A_2 = a_2\} = \frac{\Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\}}{\sum_{a_1=1}^{M_1} \Pr\{A_2 = a_2 | A_1 = a_1\} \Pr\{A_1 = a_1\}} \quad (19)$$

が得られる. 式 (17) および式 (19) がベイズの公式である. 式 (19) は $A_1 = a_1$ という事象が起こる確率 $\Pr\{A_1 = a_1\}$ と $A_1 = a_1$ という事象が起こったという条件のもとで事象 $A_2 = a_2$ が起こる確率 $\Pr\{A_2 = a_2 | A_1 = a_1\}$ から, 事象 $A_2 = a_2$ が起こったという条件のもとで事象 $A_1 = a_1$ が起こっている確率 $\Pr\{A_1 = a_1 | A_2 = a_2\}$ が表現できるということを表している. ベイズ統計では

$\Pr\{A_1 = a_1\}$ は事前確率, $\Pr\{A_1 = a_1|A_2 = a_2\}$ は事後確率と呼ばれている. ベイズ統計の戦略は, 一言で言えば, $A_1 = a_1$ は原情報, $A_2 = a_2$ はデータに対応し, 原情報が生成され, それがデータに変換されるという順過程からベイズの公式を用いて逆過程に対する確率, すなわち事後確率を構成し, これをもとにデータから原情報を推定しようというものである.

もう少し複雑な場合として, 3 つの確率変数 A_1, A_2, A_3 による場合を考えてみる. 結合確率 $\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\}$ は条件付き確率から次のように 2 つの表現で与えられる.

$$\begin{aligned}\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} &= \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_1 = a_1, A_2 = a_2\} \\ &= \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\}\end{aligned}\quad (20)$$

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} = \Pr\{A_1 = a_1, A_2 = a_2|A_3 = a_3\}\Pr\{A_3 = a_3\} \quad (21)$$

つまり,

$$\begin{aligned}\Pr\{A_1 = a_1, A_2 = a_2|A_3 = a_3\}\Pr\{A_3 = a_3\} \\ = \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\}\end{aligned}\quad (22)$$

が成り立つわけである. この両辺を a_2 に関して和をとることにより,

$$\begin{aligned}\Pr\{A_1 = a_1|A_3 = a_3\}\Pr\{A_3 = a_3\} \\ = \sum_{a_2=1}^{M_2} \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\}\end{aligned}\quad (23)$$

すなわち,

$$\begin{aligned}\Pr\{A_1 = a_1|A_3 = a_3\} \\ = \frac{1}{\Pr\{A_3 = a_3\}} \sum_{a_2=1}^{M_2} \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\}\end{aligned}\quad (24)$$

$$\Pr\{A_3 = a_3\} = \sum_{a_1=1}^{M_1} \sum_{a_2=1}^{M_2} \Pr\{A_3 = a_3|A_1 = a_1, A_2 = a_2\}\Pr\{A_2 = a_2|A_1 = a_1\}\Pr\{A_1 = a_1\} \quad (25)$$

が導かれる. 式 (24) は事象 $A_1 = a_1$ が起こり, 事象 $A_2 = a_2$ が起こり, そしてその結果として事象 $A_3 = a_3$ が起こるという順過程についての確率が与えられたときに, 逆に事象 $A_3 = a_3$ が起こったという状況のもとで事象 $A_1 = a_1$ が起こっていたかどうかという逆過程に対する条件付き確率を構成する形をとっており, その意味でベイズの公式の拡張版と見なすことができる. 一般にこのような順過程の確率から逆過程を条件付き確率と結合確率の間の関係式をもとに構成してゆく手順を総称してベイズ規則とよんでいる.

3 個の確率変数 A_1, A_2, A_3 に対するベイズ規則において結合確率 $\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\}$ が

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} = P(a_1, a_2, a_3) \quad (26)$$

という形にある関数 $P(a_1, a_2, a_3)$ を用いて与えられれば、逆過程における推論に必要な条件付き確率は定義されることを表しているわけであるが、この結合確率分布は

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} = \exp(-E(a_1, a_2, a_3)) \quad (27)$$

$$E(a_1, a_2, a_3) \equiv -\ln P(a_1, a_2, a_3) \quad (28)$$

と書き直すことができる。これは $z = \exp(\ln(z))$ というよく知られた等式を使っただけであるが、式 (27) をみると $E(a_1, a_2, a_3)$ をハミルトニアンと見ることもできる。式 (27) を形式的に

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3\} = \frac{\exp(-E(a_1, a_2, a_3))}{\sum_{z_1=1}^{M_1} \sum_{z_2=1}^{M_2} \sum_{z_3=1}^{M_3} \exp(-E(z_1, z_2, z_3))} \quad (29)$$

と書き換えれば更にわかりやすくなる。つまり、確率モデルは $P(a_1, a_2, a_3) > 0$ であれば、基本的には統計力学で言うギブス分布 (ボルツマン分布) で表せるのである。あとは $E(a_1, a_2, a_3)$ がどのような形で与えられるかで、場合によっては統計力学でよく研究されているモデルと対応がつけられることもあるわけである。この状況は確率変数の個数がいくら増えていっても可算個であるかぎり同様である。確率変数の個数、すなわち体系の大きさ (サイズ) が大きくなればそれだけ計算は大変になる。ここで統計力学が更に威力を発揮する。すなわち、統計力学が熱力学的極限という意味での巨大なサイズを持つ体系の巨視的性質を研究してきたわけであるが、その計算技術を適用することができるというわけである。

4 カルバック・ライブラー情報量と自由エネルギー

本節では、統計力学的計算手法の基礎として、自由エネルギーの解釈を情報理論におけるカルバック・ライブラー情報量という量にもとづいて説明する。

ある確率変数 A とその実現値 a に対して 2 つの結合確率分布 $P(a)$ と $Q(a)$ を考えたとき、

$$D[P||Q] \equiv \sum_z Q(z) \ln \left(\frac{Q(z)}{P(z)} \right) \quad (30)$$

という量を導入する。この量はカルバック・ライブラー情報量と呼ばれ、この 2 つの確率分布の分布間近さに対応しており、次の性質を持つ。

$$(i) D[P||Q] \geq 0$$

$$(ii) P(a) = Q(a) \Rightarrow D[P||Q] = 0$$

任意の $x > 0$ に対して $\ln(x) \leq x - 1$ が成り立ち、等号は $x = 1$ のときのみ成り立つ。このことから得られる不等式 $\ln\left(\frac{P(z)}{Q(z)}\right) \leq \frac{P(z)}{Q(z)} - 1$ を $D[P||Q]$ の表式に適用することにより、

$$D[P||Q] \equiv \sum_z Q(z) \ln \left(\frac{Q(z)}{P(z)} \right) \geq \sum_z Q(z) \left(1 - \frac{P(z)}{Q(z)} \right) \geq \sum_z P(z) - \sum_z Q(z) = 1 - 1 = 0 \quad (31)$$

という形で $D[P||Q] \geq 0$ が示される。もちろん, $D[P||Q] = D[Q||P]$ が常に成り立つわけではないので数学的意味に置いて距離と呼ぶことは言い過ぎであるが, 2つの確率分布間の近さを表す量として情報理論などでよく用いられる。

そこで \mathbf{a} のある関数 $E(\mathbf{a})$ を考え, この $E(\mathbf{a})$ に対して

$$P(\mathbf{a}) = \frac{\exp(-E(\mathbf{a}))}{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))} \quad (T > 0) \quad (32)$$

という形で与えられる確率分布を導入する。この確率分布は統計力学ではギブス分布あるいはボルツマン分布と呼ばれて, 関数 $E(\mathbf{a})$ はハミルトニアンに, T は温度に対応している。このギブス分布を式 (30) に代入すると,

$$D[P||Q] = \mathcal{F}[Q] + \ln\left(\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))\right) \quad (33)$$

$$\mathcal{F}[Q] \equiv \sum_{\mathbf{z}} E(\mathbf{z})Q(\mathbf{z}) - \left(-\sum_{\mathbf{z}} Q(\mathbf{z}) \ln(Q(\mathbf{z}))\right) \quad (34)$$

が得られる。 $\mathcal{F}[Q]$ は自由エネルギーに対応しており, 式 (34) の左辺の第 1 項は内部エネルギーに第 2 項はエントロピーに対応している。確率分布 $P(\mathbf{a})$ がギブス分布 (32) で与えられたときに, $D[P||Q]$ をできるだけ小さくするようにして, 確率分布 $Q(\mathbf{a})$ を $P(\mathbf{a})$ に近づけるということは, 自由エネルギー $\mathcal{F}[Q]$ をできるだけ小さくするように $Q(\mathbf{a})$ を選ぶことに対応していると言うことができる。

規格化条件 $\sum_{\mathbf{z}} Q(\mathbf{z}) = 1$ を拘束条件として自由エネルギー $\mathcal{F}[Q]$ の最小化の変分原理

$$\hat{Q}(\mathbf{z}) = \operatorname{argmin}_Q \left\{ \mathcal{F}[Q] \mid \sum_{\mathbf{z}} Q(\mathbf{z}) = 1 \right\} \quad (35)$$

を考えることにより, ギブス分布を逆に導くことができる。まず, 規格化条件に対してラグランジュの未定係数 λ を導入する。

$$\mathcal{L}[Q] \equiv \mathcal{F}[Q] - \lambda \left(\sum_{\mathbf{z}} Q(\mathbf{z}) - 1 \right) \quad (36)$$

汎関数 $\mathcal{L}[Q]$ を $Q(\mathbf{z})$ について変分をとると次のような極値条件が得られる。

$$\hat{Q}(\mathbf{a}) = \exp(-E(\mathbf{a}) - 1 + \lambda) \quad (37)$$

式 (37) を規格化条件の式に代入することにより λ が決定され, $\hat{Q}(\mathbf{z})$ の表式は式 (32) の右辺として与えられる。また, 式 (32) の右辺を式 (34) の $Q(\mathbf{z})$ に代入することにより

$$\mathcal{F}[\hat{Q}] = \min_Q \mathcal{F}[Q] = -\ln\left(\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))\right) \quad (38)$$

という表式として最小化された自由エネルギーの表式が得られる。

5 平均場近似の情報論的理解

本節では大規模統計モデルに対する統計力学的近似解析手法としてなじみ深い平均場近似を自由エネルギー最小の変分原理という立場から解説する。

平均場近似というと物理を専攻する学生なら、学部か大学院の講義で教わると思うが、たいていは「個々の確率変数に着目してその周りからの確率変数からの影響のある種の期待値で置き換えてしまう近似である。」と言う形に教わる。例えば、簡単のために正方格子 $\Omega = \{(x, y) | x = 1, 2, \dots, M, y = 1, 2, \dots, N\}$ 上のイジング模型を考える。各格子点 (x, y) 上の確率変数 $S_{x,y}$ はそれぞれ ± 1 の値をとり、その確率分布は

$$\Pr\{S = s\} = P(s) \equiv \frac{\exp\left(\sum_{(x,y) \in \Omega} (B_{x,y} s_{x,y} + C s_{x,y} s_{x+1,y} + C s_{x,y} s_{x,y+1})\right)}{\sum_z \exp\left(\sum_{(x,y) \in \Omega} (B_{x,y} z_{x,y} + C z_{x,y} z_{x+1,y} + C z_{x,y} z_{x,y+1})\right)} \quad (39)$$

により与えられる。ここで \sum_z は

$$\sum_z = \prod_{(x,y) \in \Omega} \sum_{z_{x,y} = \pm 1} \quad (40)$$

により定義される。平均場近似では確率変数 $S_{x,y}$ の期待値 $m_{x,y} \equiv \sum_z z_{x,y} \Pr\{S = z\}$ を用いて確率変数 S はその実現値として $(s_{x,y} - m_{x,y})(s_{x+1,y} - m_{x+1,y}) \simeq 0$ および $(s_{x,y} - m_{x,y})(s_{x,y+1} - m_{x,y+1}) \simeq 0$ を満たす $s = \{s_{x,y} | (x,y) \in \Omega\}$ の確率がそれ以外の場合に比べて非常に高いと仮定する。これがいわゆる「ゆらぎを無視する」という意味である。この場合、 $(s_{x,y} - m_{x,y})(s_{x+1,y} - m_{x+1,y}) \simeq 0$ は次のように変形される。

$$s_{x,y} s_{x+1,y} \simeq s_{x,y} m_{x+1,y} + m_{x,y} s_{x+1,y} - m_{x,y} m_{x+1,y} \quad (41)$$

$(s_{x,y} - m_{x,y})(s_{x,y+1} - m_{x,y+1}) \simeq 0$ についても同様である。これを式 (39) に代入し、 $m_{x,y}$ を計算することにより次のような $\{m_{x,y}\}$ に対する漸化式を導くことができる (図 1)。

$$m_{x,y} = \tanh\left(B_{x,y} + C(m_{x+1,y} + m_{x-1,y} + m_{x,y+1} + m_{x,y-1})\right) \quad (42)$$

同じ方程式は自由エネルギー最小の変分原理からも導出できる。この確率分布に対して各確率変数 $S_{x,y}$ ごとの周辺確率分布

$$P_{x,y}(s_{x,y}) \equiv \sum_z \delta_{s_{x,y}, z_{x,y}} P(s) \quad (43)$$

を計算する必要があるとする。定義に従って計算しようとする $2^{|\Omega|-1}$ 通りの状態についての和を計算しなければならない。そこで、この確率分布 $P(s)$ を近似する確率分布として

$$Q(s) \equiv \prod_{(x,y) \in \Omega} Q_{x,y}(s_{x,y}) \quad (44)$$

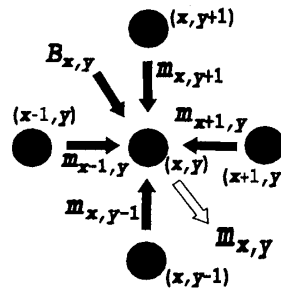


図 1: 平均場方程式 (42) の構造.

$$Q_{x,y}(s_{x,y}) \equiv \sum_z \delta_{s_{x,y}, z_{x,y}} Q(s) \quad (45)$$

を満たす分布 $Q(s)$ を導入し, 自由エネルギー (34) に代入する.

$$\begin{aligned} \mathcal{F}[Q] &= \mathcal{F}_{\text{MFA}}[\{Q_{x,y} | (x,y) \in \Omega\}] \\ &\equiv - \sum_{(x,y) \in \Omega} \left(\sum_{\zeta = \pm 1} \zeta Q_{x,y}(\zeta) \right) \left(B_{x,y} + C \sum_{\zeta = \pm 1} \zeta Q_{x+1,y}(\zeta) + C \sum_{\zeta = \pm 1} \zeta Q_{x,y+1}(\zeta) \right) \\ &\quad + \sum_{(x,y) \in \Omega} \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) \ln Q_{x,y}(\zeta) \end{aligned} \quad (46)$$

規格化条件 $\sum_{\zeta = \pm 1} Q_{x,y}(\zeta) = 1$ を拘束条件として $\{Q_{x,y}(\zeta) | (x,y) \in \Omega\}$ についての変分をとることにより, $\mathcal{D}[P||Q]$ という尺度で $Q(s)$ が $P(s)$ に最も近くなるように決めた $\{Q_{x,y}(\zeta) | (x,y) \in \Omega\}$ を $\{\hat{Q}_{x,y}(\zeta) | (x,y) \in \Omega\}$ により表す.

$$\hat{Q}_{x,y}(\zeta) = \underset{Q}{\operatorname{argmin}} \left\{ \mathcal{F}_{\text{MFA}}[\{Q_{x,y} | (x,y) \in \Omega\}] \mid \sum_{\zeta = \pm 1} Q_{x,y}(\zeta) = 1 \ ((x,y) \in \Omega) \right\} \quad (47)$$

カルバック・ライブラー情報量 $\mathcal{D}[P||Q]$ を最小にすることと, 自由エネルギー $\mathcal{F}[Q]$ を最小にすることは等価なので, $\{\hat{Q}_{x,y}(\zeta) | (x,y) \in \Omega\}$ は $\{P_{x,y}(\zeta) | (x,y) \in \Omega\}$ の良い近似になっているものと見なすことができる.

規格化条件 $\sum_{\zeta = \pm 1} Q_{x,y}(\zeta) = 1$ に対するラグランジュの未定係数 $\lambda_{x,y}$

$$\mathcal{L}[\{Q_{x,y} | (x,y) \in \Omega\}] \equiv \mathcal{F}_{\text{MFA}}[\{Q_{x,y} | (x,y) \in \Omega\}] - \sum_{(x,y) \in \Omega} \lambda_{x,y} \left(\sum_{\zeta = \pm 1} Q_{x,y}(\zeta) - 1 \right) \quad (48)$$

という形で導入し, $\{Q_{x,y}(\zeta) | (x,y) \in \Omega\}$ についての変分をとり, その後に規格化条件を満たすように $\lambda_{x,y}$ を決めることにより, $\{\hat{Q}_{x,y}(\zeta) | (x,y) \in \Omega\}$ に対する決定方程式は次のように得られる.

$$\hat{Q}_{x,y}(\zeta) = \frac{\exp \left(\left(B_{x,y} + C \sum_{(x',y') \in c_{x,y}} \left(\sum_{\zeta' = \pm 1} \zeta' \hat{Q}_{x',y'}(\zeta') \right) \right) \zeta \right)}{\sum_{\zeta = \pm 1} \exp \left(\left(B_{x,y} + C \sum_{(x',y') \in c_{x,y}} \left(\sum_{\zeta' = \pm 1} \zeta' \hat{Q}_{x',y'}(\zeta') \right) \right) \zeta \right)} \quad (49)$$

$$c_{x,y} \equiv \{(x+1, y), (x-1, y), (x, y+1), (x, y-1)\} \quad (50)$$

ここで, $S_{x,y}$ の期待値の定義を周辺確率分布 $Q_{x,y}(\zeta)$ で書き換えた表式

$$m_{x,y} \simeq \sum_z z_{x,y} Q(z) = \sum_{\zeta=\pm 1} \zeta Q_{x,y}(\zeta) \quad (51)$$

を式 (49) に代入することにより式 (42) に与えられる, 物理においてなじみのある正方格子上的イジング模型に対する平均場方程式が得られる. ひとつの見方ではあるが, 平均場近似とは与えられた確率分布をその周辺確率分布の積の形に与えられた関数系で与えられた確率分布でカルバック・ライブラー情報量という尺度のもとで最も近くなるように決めた近似であると言える.

通常, 物理で教わるイジング模型は $B_{x,y} = h/T$ ($(x,y) \in \Omega$), $C = J/T$ と置いたものであり, h は外場, J は相互作用, T は温度と呼ばれ, 格子 Ω のサイズも $M \times N$ の有限系ではなく, 無限に大きなサイズを持つ場合が考えられる. この場合, 確率変数 $S_{x,y}$ の期待値 $m_{x,y}$ は格子点 (x,y) の位置によらず一定であるから, $m_{x,y} = m$ と置くことができる. m は統計力学では秩序パラメータとよばれ, 強磁性体を想定すれば $s_{x,y}$ はスピン, h は磁場であり, m は磁化に対応する. 従って平均場方程式 (42) は

$$m = \tanh\left(\frac{h}{T} + \frac{4J}{T}m\right) \quad (52)$$

という形に与えられる. 式 (42) をみて, 「こんな式, 覚えがないぞ」と思った物理学科出身の学生さんも式 (52) をみればどこかで見たと思っていただけのものと思う. 式 (52) を h と J を固定して様々の T の値に対して反復法を用いて数値的に解くアルゴリズムを以下に与える.

平均場方程式 (52) を解く反復計算アルゴリズム

Step 1: J と h の値を入力する.

Step 2: T の初期値を $T = 8$ に設定する.

Step 3: 初期値 $m = 0$ を設定する.

Step 4: 次の操作を $\varepsilon < 10^{-6}$ となるまで繰り返す.

$$\mu \leftarrow m, \quad m \leftarrow \tanh\left(\frac{h}{T} + \frac{4J}{T}\mu\right), \quad \varepsilon \leftarrow |m - \mu|$$

Step 5: $T \leftarrow T - 0.10$ として, $T/J < 0.10$ なら終了し, そうでなければ Step 4 へ戻る.

$h = 0.00010$, $J = 1$ と設定して, 上の反復法を用いて計算した m の値を図 2 に与える. 図 2 は統計力学の教科書でよく目にするイジング模型の磁場のないときの磁化の温度依存性の曲線である.²

6 確率的画像処理のベイズ的アプローチ

ここから, 具体的情報処理の問題に入ることにする. 本節では 2 値画像の画像修復を例にとり, ごくごく簡単な場合限定して, ベイズ統計の立場で確率的画像処理の枠組みを解説する. なお, 本節および次節で与えられる定式化の詳細は文献 [6] に与えられている.

²テクニカルなことであるが, 磁場がないと称して, 実際には $h = 0.00010$ と設定しているのは式 (52) を完全に $h = 0$ と設定して, $m = 0$ を初期値として反復法で解いた場合には $m = 0$ という解しか得ることができないので, ほんの少しだけ $h = 0.00010$ という形で磁場をかけて $m \neq 0$ の解が存在する場合にはそれが得られるようにしている.

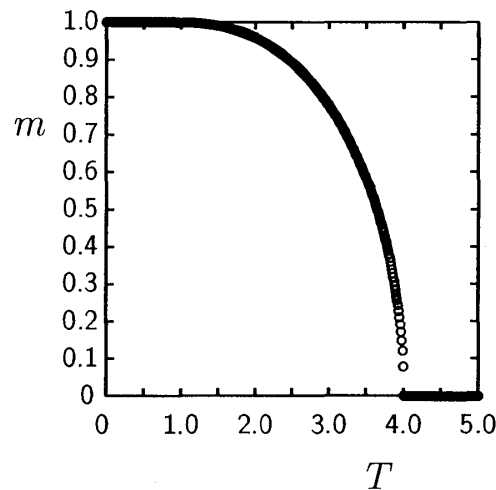


図 2: $h = 0.00010$, $J = 1$ の場合に平均場方程式 (52) を反復法により解いて得られた m の T 依存性.

まず, 画素の位置を (x, y) , 原画像および劣化画像の階調値についての確率変数を $F_{x,y}$ および $G_{x,y}$ により表すことにする. 画素は正方格子 $\Omega \equiv \{(x, y) | x = 1, 2, \dots, M, y = 1, 2, \dots, N\}$ 上に並べられ周期境界条件が課されているものとする. 2 次元的なラベル付けをされた確率変数の集合は確率場と呼ばれることがある. 原画像の確率場は $F = \{F_{x,y} | (x, y) \in \Omega\}$, 劣化画像の確率場は $g = \{g_{x,y} | (x, y) \in \Omega\}$ である. このとき, 原画像 $f = \{f_{x,y} | (x, y) \in \Omega\}$ と劣化画像 $g = \{g_{x,y} | (x, y) \in \Omega\}$ に対してベイズの公式は次のように与えられる

$$\Pr\{F = f | G = g\} = \frac{\Pr\{G = g | F = f\} \Pr\{F = f\}}{\sum_z \Pr\{G = g | F = z\} \Pr\{F = z\}} \quad (53)$$

ここで \sum_z は

$$\sum_z = \prod_{(x,y) \in \Omega} \sum_{z_{x,y} = \pm 1} \quad (54)$$

により定義される. $\Pr\{G = g | F = f\}$ は原画像 f から劣化画像 g が生成される確率である. $\Pr\{F = f\}$ は原画像 f そのものの事前確率である. ベイズの公式の右辺は, 事前確率 $\Pr\{F = f\}$ をもとに, まず原画像 f が生成され, その原画像 f から劣化過程 $\Pr\{G = g | F = f\}$ を通して劣化画像 g が生成されるといういわゆる順過程の状況を表している. 観測された劣化画像はこの順過程によって生成されたものであると考えれば, ベイズの公式は「順過程を表す式」が「劣化画像 g が与えられたという条件のもとでの原画像 f に対する事後確率 $\Pr\{F = f | G = g\}$ 」に等しいということの意味している (図 3).

最も基本的な場合として, 各画素が階調値として -1 と 1 のみをとる 2 値画像を考え, 「劣化過程は原画像から各画素で独立にある確率で劣化されていること」, 「原画像の各画素の階調値はその近傍画素の階調値と同じ値をとる確率が高い」という 2 つの仮定が基本的であると考え, これらを明確に式を用いて書き下すと次のようになる.

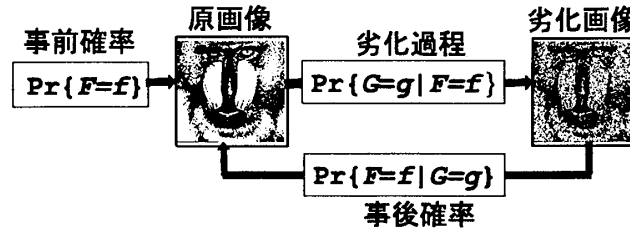


図 3: ベイズ統計における原画像の推定メカニズム.

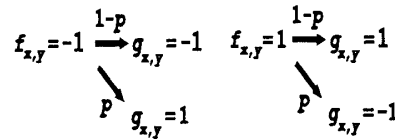


図 4: 2 元対称通信路.

劣化過程 劣化画像 g は原画像 f の各画素の状態が各画素ごとに独立に確率 p で -1 から 1 へ、または 1 から -1 へと置き換わることにより生成されるものとする (図 4).

$$\Pr\{G = g|F = f\} \equiv \prod_{(x,y) \in \Omega} \left((1-p)\delta_{f_{x,y}, g_{x,y}} + p(1 - \delta_{f_{x,y}, g_{x,y}}) \right) \quad (55)$$

原画像に対する事前情報 原画像 f は次の確率分布の高い確率を与える画像のひとつである (図 5).

$$\begin{aligned} & \Pr\{F = f\} \\ & \equiv \frac{\left(\prod_{(x,y) \in \Omega} \exp\left(-\frac{1}{2}\alpha(f_{x,y} - f_{x+1,y})^2\right) \right) \left(\prod_{(x,y) \in \Omega} \exp\left(-\frac{1}{2}\alpha(f_{x,y} - f_{x,y+1})^2\right) \right)}{\sum_z \left(\prod_{(x,y) \in \Omega} \exp\left(-\frac{1}{2}\alpha(z_{x,y} - z_{x+1,y})^2\right) \right) \left(\prod_{(x,y) \in \Omega} \exp\left(-\frac{1}{2}\alpha(z_{x,y} - z_{x,y+1})^2\right) \right)} \end{aligned} \quad (56)$$

確率変数 $f_{x,y}$ の集合すなわち確率場 f は各画素 (x, y) の階調値 $f_{x,y}$ がその近傍画素 $(x \pm 1, y)$, $(x, y \pm 1)$ の階調値にのみ依存する形になっており、この性質を持つ確率場 f は総称してマルコフ確率場と呼ばれる.

2 元対称通信路は情報理論で考えられる最も基本的な劣化過程の一つである. 図 4 に示すように各画素で独立に原画像と異なる階調値に置き換えられるということなので, 例えば $p = 0.2$ と設定すると原画像と劣化画像で 20% 程度の画素が劣化されているということになる. 式 (56) は最近接画素対が同じ階調値をとればとるほど確率が高くなるわけなので, 当然, 最大確率を与える画像は真っ白か真っ黒な画像である. そんな画像ばかり扱っていたのでは使えない. 当然の話である. しかし, どの程度かなどという堅いことを言わないで比較的高い確率を与える画像はと考えると話は多少変わってくる. いくつかの α の値に対してモンテカルロシミュレーションのスナップショットとして生成され

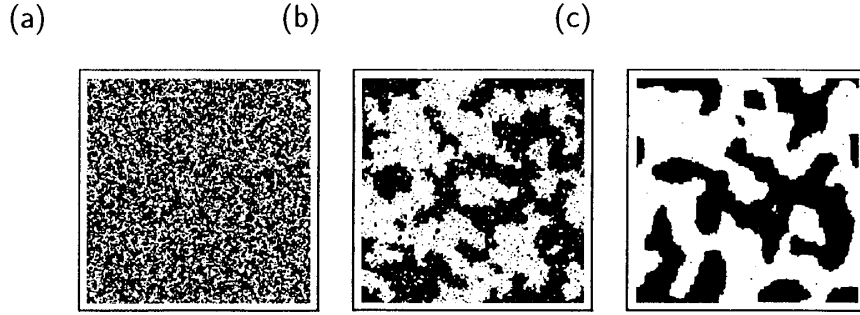


図 5: いくつかの α の値に対して確率分布 (56) に従うモンテカルロ・シミュレーションのスナップショットとして生成された 2 値画像. (a) $\alpha = 0.25$. (b) $\alpha = 0.5$. (c) $\alpha = 1$.

た画像の例が図 5 である. 確率モデル (56) のスナップショットを持ってきてこれがわれわれの扱える原画像でございましてと言われても現実世界の画像は様々な意味と構造を持っていて, そんな画像で置き換えられるわけではない. しかし, このような仮定の下で画像処理アルゴリズムを構成すると現実世界の画像も含めてうまく処理できてしまうから不思議である. 著者がこの説明をするとき, 「現実世界の画像の部分部分のパターンをみると図 5 のような画像がみえてくると言うことは無理がないとは考えられませんか?」という苦しい説明をする. このことを深く深く考えてゆくと話がここで終わってしまうので, これについてはとりあえず認めていただくということで話をすすめていきたい.

式 (55) および式 (56) をベイズの公式 (53) に代入することにより, $\Pr\{F = f|G = g\}$ は次のように与えられる.

$$\Pr\{F = f|G = g\} = \frac{\exp(-E(f|g))}{\sum_z \exp(-E(z|g))} \quad (57)$$

$$E(f|g) \equiv - \sum_{(x,y) \in \Omega} \left(-\frac{1}{2}\beta(g_{x,y} - f_{x,y})^2 - \frac{1}{2}\alpha(f_{x,y} - f_{x+1,y})^2 - \frac{1}{2}\alpha(f_{x,y} - f_{x,y+1})^2 \right) \quad (58)$$

$$\beta \equiv \frac{1}{2} \ln \left(\frac{1-p}{p} \right) \quad (59)$$

修復画像 $\hat{f} = \{\hat{f}_{x,y} | (x,y) \in \Omega\}$ はこの事後確率分布 $\Pr\{F = f|G = g\}$ から

$$\hat{f}_{x,y} = \operatorname{argmax}_{f_{x,y}} \Pr\{F_{x,y} = f_{x,y} | G = g\} \quad (60)$$

$$\Pr\{F_{x,y} = f_{x,y} | G = g\} \equiv \sum_z \delta_{f_{x,y}, z_{x,y}} \Pr\{F = z | G = g\} \quad (61)$$

により, 各画素ごとに階調値を推定することにより決定される. $\Pr\{F_{x,y} = f_{x,y} | G = g\}$ は事後確率分布 $\Pr\{F = f | G = g\}$ の確率変数 $f_{x,y}$ についての周辺確率分布とよばれる. 式 (61) の定義にもとづいて $\Pr\{F_{x,y} = f_{x,y} | G = g\}$ を計算しようとする, $2^{|\Omega|-1}$ 通りの画像についての和を計算する必要があり, 膨大な計算量が必要となる. 次節ではこれを計算する方法として確率伝搬法について説明する.

7 ベーテ近似を用いた確率的画像処理アルゴリズム

本節では、確率モデル (57) の各画素ごとの確率変数 $F_{x,y}$ の周辺確率分布 $\Pr\{F_{x,y} = f_{x,y} | G = g\}$ を高精度に計算する近似として、ベーテ近似について自由エネルギー最小の変分原理にもとづいて概説する。ベーテ近似は通常、平均場近似の拡張版としての統計力学では大学院レベルの講義で顔をだすことがあるので、本講義を受講されている学生諸氏はほとんどの方が初めてであろう。

まず、第3節の説明と同様にして式 (57) で与えられるギブス分布は次のような自由エネルギー最小の変分原理を満足する。

$$\Pr\{F = f | G = g\} = \min_Q \mathcal{F}[Q] \quad (62)$$

$$\mathcal{F}[Q] \equiv \mathcal{E}[Q] - \mathcal{S}[Q] \quad (63)$$

$$\mathcal{E}[Q] \equiv \sum_z E(z|g)Q(z), \quad \mathcal{S}[Q] \equiv - \sum_z Q(z) \ln Q(z) \quad (64)$$

ここで、試行関数 $Q(f)$ に対して

$$Q_{x,y}(f_{x,y}) \equiv \sum_z \delta_{f_{x,y}, z_{x,y}} Q(z) \quad (65)$$

$$Q_{x,y}^{x',y'}(f_{x,y}, f_{x',y'}) \equiv \sum_z \delta_{f_{x,y}, z_{x,y}} \delta_{f_{x',y'}, z_{x',y'}} Q(z) \quad (66)$$

を導入する。 $Q_{x,y}(f_{x,y})$ および $Q_{x,y}^{x',y'}(f_{x,y}, f_{x',y'})$ は $f(s) = P(s|h)$ であれば $\Pr\{F = f | G = g\}$ のある画素または画素対の周辺確率分布であることは容易に理解できる。式 (58), 式 (65) および式 (66) を式 (64) に代入することにより

$$\begin{aligned} \mathcal{E}[Q] = & - \sum_{(x,y) \in \Omega} \left(\frac{1}{2} \beta \sum_{\zeta = \pm 1} (\zeta - g_{x,y})^2 Q_{x,y}(\zeta) \right. \\ & \left. + \frac{1}{2} \alpha \sum_{\zeta = \pm 1} \sum_{\zeta' = \pm 1} (\zeta - \zeta')^2 (Q_{x,y}^{x+1,y}(\zeta, \zeta') + Q_{x,y}^{x,y+1}(\zeta, \zeta')) \right) \end{aligned} \quad (67)$$

という式が導かれる。すなわち、これで汎関数 $\mathcal{E}[Q]$ は周辺確率分布 $Q_{x,y}(f_{x,y})$, $Q_{x,y}^{x+1,y}(f_{x,y}, f_{x+1,y})$ および $Q_{x,y}^{x,y+1}(f_{x,y}, f_{x,y+1})$ のみにより表されたことになる。ここで、 $\mathcal{S}[Q]$ についても同様の書き換えができれば、 $\mathcal{F}[Q]$ も周辺確率分布 $Q_{x,y}(f_{x,y})$, $Q_{x,y}^{x+1,y}(f_{x,y}, f_{x+1,y})$ および $Q_{x,y}^{x,y+1}(f_{x,y}, f_{x,y+1})$ のみにより表されたことになるが、厳密にそのような書き換えができると考えることは難しい。式 (57) で与えられた確率分布は次の形に書き換えることができる。

$$\begin{aligned} \Pr\{F = f | G = g\} = & \frac{1}{Z} \left(\prod_{(x,y) \in \Omega} W_{x,y}(f_{x,y}) \right) \\ & \times \left(\prod_{(x,y) \in \Omega} \frac{W_{x,y}^{x+1,y}(f_{x,y}, f_{x+1,y})}{W_{x,y}(f_{x,y}) W_{x+1,y}(f_{x+1,y})} \right) \left(\prod_{(x,y) \in \Omega} \frac{W_{x,y}^{x,y+1}(f_{x,y}, f_{x,y+1})}{W_{x,y}(f_{x,y}) W_{x,y+1}(f_{x,y+1})} \right) \end{aligned} \quad (68)$$

ここで Z は規格化定数であり, $W_{x,y}(\xi)$ および $W_{x,y}^{x',y'}(\xi, \xi')$ は次のように定義される.

$$W_{x,y}(\xi) \equiv \frac{\exp\left(-\frac{1}{2}\beta(\xi - g_{x,y})^2\right)}{\sum_{\zeta=\pm 1} \exp\left(-\frac{1}{2}\beta(\zeta - g_{x,y})^2\right)} \quad (69)$$

$$W_{x,y}^{x',y'}(\xi, \xi') \equiv \frac{\exp\left(-\frac{1}{2}\beta(\xi - g_{x,y})^2 - \frac{1}{2}\beta(\xi' - g_{x',y'})^2 - \frac{1}{2}\alpha(\xi - \xi')^2\right)}{\sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} \exp\left(-\frac{1}{2}\beta(\zeta - g_{x,y})^2 - \frac{1}{2}\beta(\zeta' - g_{x',y'})^2 - \frac{1}{2}\alpha(\zeta - \zeta')^2\right)} \quad (70)$$

式 (68) の形をもとにベータ近似では $Q(f)$ をその周辺確率分布を用いて

$$Q(f) = \left(\prod_{(x,y) \in \Omega} Q_{x,y}(f_{x,y}) \right) \left(\prod_{(x,y) \in \Omega} \frac{Q_{x,y}^{x+1,y}(f_{x,y}, f_{x+1,y})}{Q_{x,y}(f_{x,y}) Q_{x+1,y}(f_{x+1,y})} \right) \\ \times \left(\prod_{(x,y) \in \Omega} \frac{Q_{x,y}^{x,y+1}(f_{x,y}, f_{x,y+1})}{Q_{x,y}(f_{x,y}) Q_{x,y+1}(f_{x,y+1})} \right) \quad (71)$$

という形に表された関数系に制限してその範囲で自由エネルギーを最小にするという尺度をもって試行関数 $Q(f)$ ができるだけ式 (57) で与えられたギブス分布に近くなるように周辺確率分布を決定するという戦略がとられている. 式 (71) を式 (64) に代入することにより, $S[Q]$ を次のような近似的表式で置き換えることができる.

$$S[Q] \simeq \sum_{(x,y) \in \Omega} \mathcal{S}_{x,y} + \sum_{(x,y) \in \Omega} \left(\mathcal{S}_{x,y}^{x+1,y} - \mathcal{S}_{x,y} - \mathcal{S}_{x+1,y} \right) + \sum_{(x,y) \in \Omega} \left(\mathcal{S}_{x,y}^{x,y+1} - \mathcal{S}_{x,y} - \mathcal{S}_{x,y+1} \right) \quad (72)$$

$$\mathcal{S}_{x,y} \equiv - \sum_{\zeta=\pm 1} Q_{x,y}(\zeta) \ln(Q_{x,y}(\zeta)) \quad (73)$$

$$\mathcal{S}_{x,y}^{x',y'} \equiv - \sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} Q_{x,y}^{x',y'}(\zeta, \zeta') \ln(Q_{x,y}^{x',y'}(\zeta, \zeta')) \quad (74)$$

この表式はまず $S[Q]$ をより小数の画素からなる周辺確率分布を用いて表そうとすると, まず現れるのが各画素ごとの $\mathcal{S}_{x,y}$ の和であると考えことは自然であろう. もし $\alpha = 0$ の場合を考えているのであればこの項のみで終わりである. ところが今考えている確率モデルでは最近接画素対間の相互作用 α が存在している. そこで最近接画素対についての周辺確率分布による項をこれに何らかの形で加える必要がある. 上式の場合, その最近接画素対からの寄与の項を, すべての最近接画素対から画素の寄与を差し引いた量 $\mathcal{S}_{x,y}^{x+1,y} - \mathcal{S}_{x,y} - \mathcal{S}_{x+1,y}$ および $\mathcal{S}_{x,y}^{x,y+1} - \mathcal{S}_{x,y} - \mathcal{S}_{x,y+1}$ として考慮している. 画素の寄与をわざわざ引いているのは第 1 項で既に足しているから差し引かなければ足しすぎになるという考え方からくるものである. 式 (117) と式 (120) から $\mathcal{F}[f]$ の近似表現は次のように与えられる.

$$\mathcal{F}[Q] \simeq \mathcal{F}_{\text{Bethe}}[\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}] \\ \equiv \sum_{(x,y) \in \Omega} \left(-\frac{1}{2}\beta \sum_{\zeta=\pm 1} (\zeta - g_{x,y})^2 Q_{x,y}(\zeta) - \frac{1}{2}\alpha \sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} (\zeta - \zeta')^2 (Q_{x,y}^{x+1,y}(\zeta, \zeta') + Q_{x,y}^{x,y+1}(\zeta, \zeta')) \right) \\ + \mathcal{S}_{x,y} + (\mathcal{S}_{x,y}^{x+1,y} - \mathcal{S}_{x,y} - \mathcal{S}_{x+1,y}) + (\mathcal{S}_{x,y}^{x,y+1} - \mathcal{S}_{x,y} - \mathcal{S}_{x,y+1}) \quad (75)$$

周辺確率分布の定義 (65), (66) から次の等式が導かれる.

$$\begin{aligned} Q_{x,y}(\zeta) &= \sum_{\zeta'=\pm 1} Q_{x,y}^{x+1,y}(\zeta, \zeta') = \sum_{\zeta'=\pm 1} Q_{x-1,y}^{x,y}(\zeta', \zeta) \\ &= \sum_{\zeta'=\pm 1} Q_{x,y}^{x,y+1}(\zeta, \zeta') = \sum_{\zeta'=\pm 1} Q_{x,y-1}^{x,y}(\zeta', \zeta) \quad (\zeta = \pm 1) \end{aligned} \quad (76)$$

そこで, $Q_{x,y}(s_{x,y})$, $Q_{x,y}^{x+1,y}(\zeta, \zeta')$, $Q_{x,y}^{x,y+1}(\zeta, \zeta')$ はその規格化条件および等式 (76) を拘束条件として $\mathcal{F}_{\text{Bethe}}[\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}]$ の最小化に対する変分原理をとることにより決定される.

$$\begin{aligned} &\{\hat{Q}_{x,y}, \hat{Q}_{x,y}^{x+1,y}, \hat{Q}_{x,y}^{x,y+1} | (x,y) \in \Omega\} \\ &= \arg \min_{\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}} \left\{ \mathcal{F}_{\text{Bethe}}[\{Q_{x,y}, Q_{x,y}^{x+1,y}, Q_{x,y}^{x,y+1} | (x,y) \in \Omega\}] \right\} \\ &\quad \begin{aligned} Q_{x,y}(f_{x,y}) &= \sum_{\zeta=\pm 1} Q_{x,y}^{x+1,y}(f_{x,y}, \zeta) = \sum_{\zeta=\pm 1} Q_{x-1,y}^{x,y}(\zeta, f_{x,y}) \\ &= \sum_{\zeta=\pm 1} Q_{x,y}^{x,y+1}(f_{x,y}, \zeta) = \sum_{\zeta=\pm 1} Q_{x,y-1}^{x,y}(\zeta, f_{x,y}), \\ \sum_{\zeta=\pm 1} Q_{x,y}(\zeta) &= \sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} Q_{x,y}^{x+1,y}(\zeta, \zeta') = \sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} Q_{x,y}^{x,y+1}(\zeta, \zeta') = 1 \end{aligned} \end{aligned} \quad (77)$$

条件付き変分原理における極値条件から $Q_{x,y}(s_{x,y})$, $Q_{x,y}^{x+1,y}(\zeta, \zeta')$, $Q_{x,y}^{x,y+1}(\zeta, \zeta')$ に対する表式が拘束条件に対して導入されたラグランジュの未定係数を用いて表され, そのラグランジュの未定係数は周辺確率分布についての規格化条件および等式 (76) を満たすように決定されるという形の非線形方程式に帰着される. このラグランジュの未定係数に対する多少の変数変換を行い, 見やすい形に整理した方程式を以下の示す.

$$Q_{x,y}(\xi) = \frac{W_{x,y}(\xi) \prod_{(x',y') \in c_{x,y}} \mathcal{M}_{x,y}^{x',y'}(\xi)}{\sum_{\zeta=\pm 1} W_{x,y}(\zeta) \prod_{(x',y') \in c_{x,y}} \mathcal{M}_{x,y}^{x',y'}(\zeta)}, \quad (78)$$

$$Q_{x,y}^{x',y'}(\xi, \xi') = \frac{W_{x,y}^{x',y'}(\xi, \xi') \prod_{(x'',y'') \in c_{x,y} \setminus (x',y')} \mathcal{M}_{x,y}^{x'',y''}(\xi) \prod_{(x''',y''') \in c_{x',y'} \setminus (x,y)} \mathcal{M}_{x',y'}^{x''',y'''}(\xi')}{\sum_{\zeta=\pm 1} \sum_{\zeta'=\pm 1} W_{x,y}^{x',y'}(\zeta, \zeta') \prod_{(x'',y'') \in c_{x,y} \setminus (x',y')} \mathcal{M}_{x,y}^{x'',y''}(\zeta) \prod_{(x''',y''') \in c_{x',y'} \setminus (x,y)} \mathcal{M}_{x',y'}^{x''',y'''}(\zeta')}, \quad (79)$$

$$\mathcal{M}_{x,y}^{x',y'}(\xi) = \frac{\sum_{\zeta=\pm 1} \left(\frac{W_{x,y}^{x',y'}(\xi, \zeta)}{W_{x,y}(\xi)} \right) \prod_{(x'',y'') \in c_{x',y'} \setminus (x,y)} \mathcal{M}_{x',y'}^{x'',y''}(\zeta)}{\sum_{\zeta'=\pm 1} \sum_{\zeta=\pm 1} \left(\frac{W_{x,y}^{x',y'}(\zeta', \zeta)}{W_{x,y}(\zeta')} \right) \prod_{(x'',y'') \in c_{x',y'} \setminus (x,y)} \mathcal{M}_{x',y'}^{x'',y''}(\zeta)}. \quad (80)$$

$$c_{x,y} \equiv \{(x+1, y), (x-1, y), (x, y+1), (x, y-1)\} \quad (81)$$

これらの式は一見複雑に見えるが実はよく見るときれいな構造を持っている。まず式 (78) で与えられた周辺確率分布は画素 (x, y) にその隣接画素からの影響を $\lambda_{x',y'}^{x,y}$ ($(x', y') \in c_{x,y}$) という形で取り込んだ構造をしている。これをグラフ表現にしたのが図 6 である。式 (79) で与えられた周辺確率分布は例えば $f_{x,y}^{x+1,y}(\zeta, \zeta')$ の場合、画素 (x, y) には $(x+1, y)$ を除く隣接画素からの影響を $\lambda_{x',y'}^{x,y}$ ($(x', y') \in c_{x,y} \setminus (x+1, y)$) という形で取り込んだ構造をしている。また、画素 $(x+1, y)$ には (x, y) を除く隣接画素からの影響を $\lambda_{x',y'}^{x,y}$ ($(x', y') \in c_{x,y} \setminus (x, y)$) という形で取り込んだ構造をしている。これをグラフ表現にしたのが図 7 である。そして、式 (83) は例えば $(x', y') = (x+1, y)$ と設定した場合、画素 (x, y) に $(x+1, y)$ を除く隣接画素から $\lambda_{x',y'}^{x,y}$ ($(x', y') \in c_{x,y} \setminus (x+1, y)$) が入力として入り、 $(x+1, y)$ へと $\lambda_{x',y'}^{x,y}$ が出力として伝搬されてゆく構造を持っている。これをグラフ表現にしたのが図 8 である。

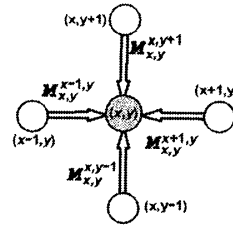


図 6: 式 (78) で与えられた周辺確率分布 $f_{x,y}(\zeta)$ のグラフ表現。

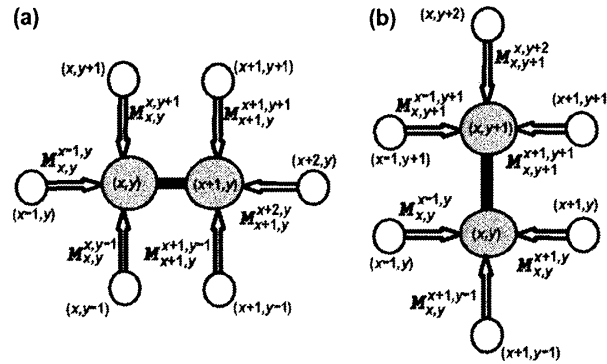


図 7: 式 (79) で与えられた周辺確率分布 $f_{x,y}^{x+1,y}(\zeta, \zeta')$ および $f_{x,y}^{x,y+1}(\zeta, \zeta')$ のグラフ表現。 (a) $(x', y') = (x+1, y)$. (b) $(x', y') = (x, y+1)$.

式 (65)-(125) において、 $\mathcal{M}_{x',y'}^{x'',y''}(\zeta)$ の変数 ζ は 0, 1 しかとらないことを考慮すると

$$\mathcal{M}_{x',y'}^{x'',y''}(-1) = \frac{\exp(-\lambda_{x',y'}^{x,y})}{2\cosh(\lambda_{x',y'}^{x,y})}, \quad \mathcal{M}_{x',y'}^{x'',y''}(1) = \frac{\exp(\lambda_{x',y'}^{x,y})}{2\cosh(\lambda_{x',y'}^{x,y})} \quad (82)$$

と変数変換することにより、特に式 (125) は次のように書き換えられる。

$$\lambda_{x',y'}^{x,y} = \operatorname{arctanh} \left(\tanh(\alpha) \tanh \left(\beta h_{x,y} + \sum_{(x'', y'') \in c_{x,y} \setminus (x', y')} \lambda_{x,y}^{x'', y''} \right) \right) \quad (83)$$

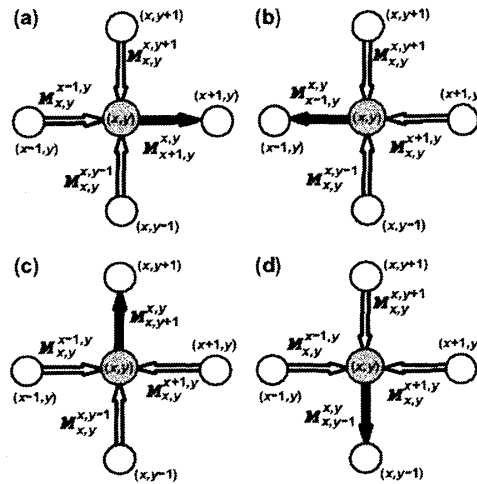


図 8: 式 (83) で与えられた反復方程式のグラフ表現. (a) $(x', y') = (x+1, y)$. (b) $(x', y') = (x-1, y)$. (c) $(x', y') = (x, y+1)$. (d) $(x', y') = (x, y-1)$.

本節で紹介したベーテ近似は式 (120) が重要な鍵になっている. そこでは最近接画素対の相関の効果までを重要視し, それより大きな画素の集合の効果は小さいとして無視している.

修復画像 $\hat{f} = \{\hat{f}_{x,y} | (x, y) \in \Omega\}$ は式 (60) により決定される. 周辺確率分布 $\Pr\{F_{x,y} = f_{x,y} | G = g\}$ は式 (78)-(80) を解き, 画素 (x, y) のおける周辺確率分布 $Q_{x,y}(f_{x,y})$ を得ることにより近似的に計算される. α と p の値をひとつ固定したときのベーテ近似を用いた画像修復のための反復計算アルゴリズムは以下に与えられる.

ベーテ近似を用いた画像修復アルゴリズム

Step 1: 原画像 $f = \{f_{x,y} | (x, y) \in \Omega\}$, 劣化画像 $g = \{g_{x,y} | (x, y) \in \Omega\}$ と p, α, R, Δ の値を設定する.

Step 2: 初期値を $\lambda_{x,y}^{x',y'} = 0$ ($(x, y) \in \Omega, (x', y') \in c_{x,y}$), $T \leftarrow 1 + R\Delta$ と設定する.

Step 3: $T \leftarrow T - \Delta, r \leftarrow 0$ と設定する.

Step 4: $r \leftarrow r + 1, a_{x,y}^{x+1,y} \leftarrow \lambda_{x,y}^{x+1,y}, a_{x,y}^{x,y+1} \leftarrow \lambda_{x,y}^{x,y+1}$ ($(x, y) \in \Omega$), と設定する.

Step 5: $\{\lambda_{x,y}^{x+1,y}, \lambda_{x,y}^{x,y+1} | (x, y) \in \Omega\}$ の値を次の更新則により更新する.

$$\mathcal{M}_{x,y}^{x',y'}(\xi) \leftarrow \frac{\sum_{\zeta=\pm 1} \left(\frac{W_{x,y}^{x',y'}(\xi, \zeta)}{W_{x,y}(\xi)} \right)^{1/T} \prod_{(x'', y'') \in c_{x',y'} \setminus (x, y)} a_{x',y'}^{x'',y''}(\zeta)}{\sum_{\zeta'=\pm 1} \sum_{\zeta=\pm 1} \left(\frac{W_{x,y}^{x',y'}(\zeta', \zeta)}{W_{x,y}(\zeta')} \right)^{1/T} \prod_{(x'', y'') \in c_{x',y'} \setminus (x, y)} a_{x',y'}^{x'',y''}(\zeta)} \quad ((x', y') \in c_{x,y}) \quad (84)$$

Step 6:

$$\frac{1}{MN} \sum_{(x,y) \in \Omega} \sum_{\zeta=\pm 1} \left(|a_{x,y}^{x+1,y}(\zeta) - \mathcal{M}_{x,y}^{x+1,y}(\zeta)| + |a_{x,y}^{x,y+1}(\zeta) - \mathcal{M}_{x,y}^{x,y+1}(\zeta)| \right) < 10^{-6} \quad (85)$$

を満足しなければ Step 4 にもどり, 満足すれば Step 7 に進む.

Step 7: $T = 1$ を満足しなければ Step 3 にもどり, 満足すれば修復画像 $\hat{f} = \{\hat{f}_{x,y} | (x, y) \in \Omega\}$ を

$$\hat{f}_{x,y} \leftarrow \arg \max_{\zeta=\pm 1} \left(W_{x,y}(\xi) \prod_{(x', y') \in c_{x,y}} \mathcal{M}_{x,y}^{x',y'}(\xi) \right) \quad (86)$$

により決定し, 終了する.

数値実験を行った一例を図 9 及び図 10 に与える. 図 9(a) の原画像は $\alpha = 0.5$ と設定し, 確率分布 (56) に従うモンテカルロ・シミュレーションにより生成された 2 値画像である. すなわち図 9 の数値実験は前節で与えたベイズ的アプローチの枠組みの順過程に忠実に従って生成されたデータである劣化画像からの原画像の推定を行った結果である. 修復の際には α は原画像を生成したときの値を用い, p も劣化させたときに設定した値をそのまま用いている. 図 10(a) の原画像は信号処理の分野で標準的に用いられている “Mandrill” と呼ばれる画像に対してある種のしきい値処理を行って作成した画像である. すなわち, 図 10 は現実世界の画像をもとに生成された 2 値画像に対して行った数値実験であり, 必ずしもデータがベイズ的アプローチの枠組みの順過程に従って生成されているとは限らない場合である. 修復の際には p は劣化させたときに設定した値をそのまま用いているが, α は, そもそも原画像自体が式 (56) を使って生成されたものではないため, 手動で修復画像が原画像に近くなるように調整して決めている. いずれの場合も良好な結果が得られることがわかる.

ここまでの話の中で p と α を劣化画像から推定することについては述べて来なかったが, 実際の画像処理システムにおいてはこの p および α を与えられたデータのみから推定しなければならない場合が多い. 原画像が式 (56) における α のどのようなモデルパラメータ α と p の推定値 $\hat{\alpha}$, \hat{p} はベイズ統計においてはハイバパラメータと呼ばれ, 周辺尤度最大化とよばれる最尤推定の立場において次のように決定されることが多い [6].

$$(\hat{\alpha}, \hat{p}) = \max_{(\alpha, p)} \sum_z \Pr\{G = g | F = z\} \Pr\{F = z\} \quad (87)$$

右辺の量 $\Pr\{G = g\} \equiv \sum_z \Pr\{G = g | F = z\} \Pr\{F = z\}$ が周辺尤度と呼ばれる量であり, これもベータ近似を用いて計算することができる.

本節と前節で与えた枠組みは各画素の階調値が 0, 1 ではなく, 任意の実数値をとる場合にも容易に拡張できる. 例えば劣化過程として平均 0, 分散 $\sigma^2 = \frac{1}{2\beta}$ の加法的白色ガウス雑音 $\mathcal{N}[0, \sigma^2]$ を考える. この場合, 劣化過程は

$$\Pr\{G = g | F = f\} \equiv \prod_{(x,y) \in \Omega} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2\sigma^2}(f_{x,y} - g_{x,y})^2\right) \quad (88)$$

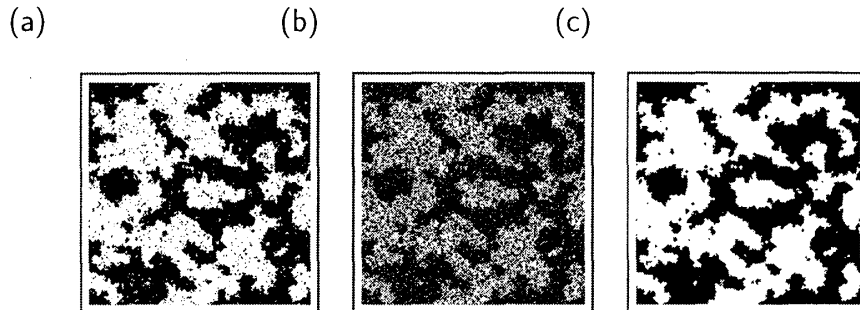


図 9: 2 値画像の画像修復例 (a) 原画像 ($\alpha = 0.5$ と設定し, 確率分布 (56) に従うモンテカルロ・シミュレーションにより生成された 2 値画像). (b) 劣化画像 ($p = 0.2$ と設定し, 劣化過程 (55) を通して, 生成された 2 値画像). (c) ベータ近似による修復画像 ($p = 0.2$, $\alpha = 0.5$)

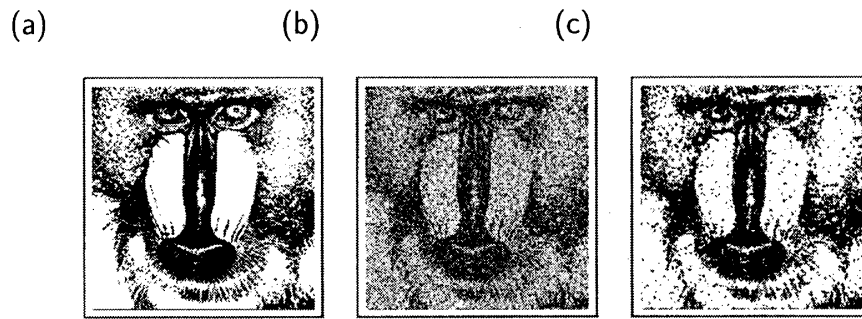


図 10: 2 値画像の画像修復例 (a) 原画像. (b) 劣化画像 ($p = 0.2$). (c) ベーテ近似による修復画像 ($p = 0.2, \alpha = 0.35$)

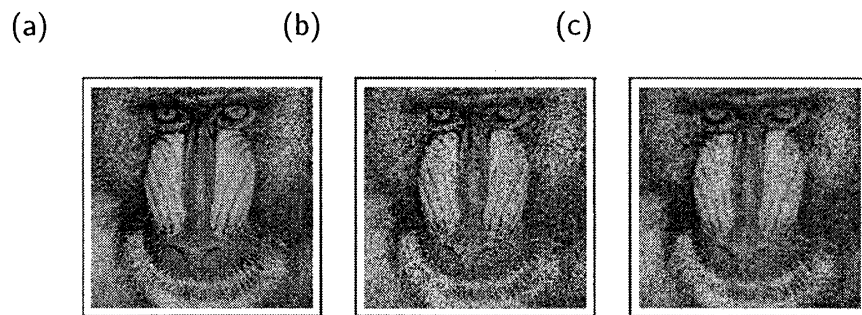


図 11: 濃淡画像の画像修復例 (a) 原画像. (b) 加法的白色ガウス雑音による劣化画像 ($\mathcal{N}[0, 30^2]$). (c) ベーテ近似による修復画像 ($\sigma = 30, \alpha = 0.0005$)

という形に与えられる. 計算の詳細は省略するが $\sigma = 20$ の場合の画像修復例を図 11 に与える.

本節で紹介した 2 値画像の画像処理アルゴリズムの C 言語を用いた具体的なプログラムはホームページ (URL: <http://www.statp.is.tohoku.ac.jp/~kazu/SMAPIP-KazuKazu/>) からダウンロードすることができる.

8 確率的推論

本節では統計力学を用いた確率的情報処理のもう一つの例として人工知能における確率的推論をとりあげる. ベイズ統計にもとづいて構成された推論機構は一般にベイジアンネットと呼ばれている. 確率推論の代表的な教科書としては Pearl の教科書等があげられる [23, 24]. また, 最近の発展については文献 [9] 等を参照していただきたい. ここで, 統計力学的手法の一つとして知られるクラスター変分法を用いた確率的推論アルゴリズムについて解説する. クラスター変分法は前節のベーテ近似を更に一般的に拡張した統計力学的手法である.

確率的推論は複数のノードと呼ばれる点とそのノード間の因果関係を表すリンクからグラフ表現の形で与えられる. そこでまず, 例として 4 個のノードを考え, 図 12 に示すようなグラフ表現で与えられる確率的推論機構を用いて, どのようなものかを初心者にわかりやすく解説しよう. この例題は K. Murphy 氏によって作成されたホームページ <http://www.ai.mit.edu/~murphyk/Bayes/bayes.html> に

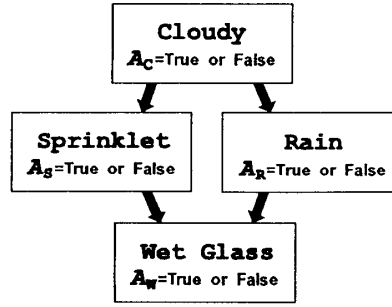


図 12: 4 個のノードからなる確率的推論機構の例.

ベイジアンネットの解説とともに掲載されているものである. その 4 個のノードに対してそれぞれに [曇りである], [スプリンクラーが作動している], [雨が降る], [芝生がぬれている] という命題を割り当て, そのそれぞれの命題に対応して true (T) と false (F) という 2 つの状態をとる確率変数をそれぞれ A_C , A_S , A_R , A_W により表すものとする. 結合確率と条件付き確率の関係式により

$$\begin{aligned}
 & \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \\
 &= \Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\} \Pr\{A_C = a_C, A_S = a_S, A_R = a_R\} \\
 &= \Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\} \Pr\{A_R = a_R | A_C = a_C, A_S = a_S\} \\
 &\quad \times \Pr\{A_C = a_C, A_S = a_S\} \\
 &= \Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\} \Pr\{A_R = a_R | A_C = a_C, A_S = a_S\} \\
 &\quad \times \Pr\{A_S = a_S | A_C = a_C\} \Pr\{A_C = a_C\}
 \end{aligned} \tag{89}$$

が導かれる. ここで, 各命題に対する因果関係として「曇りである」とすればいずれ「雨が降る」可能性がある. 「曇りでない」とすれば晴れているということで芝生を枯らさないために「スプリンクラーを作動させる」ということになる. そして「雨が降る」かまたは「スプリンクラーを作動させる」ということにより「芝生がぬれている」という状況が発生する. $\Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\}$ は A_C すなわち「曇っているかどうか」には直接は依存せず, $\Pr\{A_R = a_R | A_C = a_C, A_S = a_S\}$ は S すなわち「スプリンクラーが作動しているかどうか」には直接依存しないことを考慮するとそれぞれ

$$\begin{aligned}
 & \Pr\{A_W = a_W | A_C = a_C, A_S = a_S, A_R = a_R\} = \Pr\{A_W = a_W | A_S = a_S, A_R = a_R\} \\
 & \Pr\{A_R = a_R | A_C = a_C, A_S = a_S\} = \Pr\{A_R = a_R | A_C = a_C\}
 \end{aligned} \tag{90}$$

と与えられる. すなわち, 結合確率 $\Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\}$ は次のように与えられる.

$$\begin{aligned}
 & \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \\
 &= \Pr\{A_W = a_W | A_S = a_S, A_R = a_R\} \Pr\{A_R = a_R | A_C = a_C\} \\
 &\quad \times \Pr\{A_S = a_S | A_C = a_C\} \Pr\{A_C = a_C\}
 \end{aligned} \tag{91}$$

表 1: 式 (91) における $\Pr\{A_W = a_A | A_S = a_S, A_R = a_R\}$, $\Pr\{A_R = a_R | A_C = a_C\}$, $\Pr\{A_S = a_S | A_C = a_C\}$, $\Pr\{A_C = a_C\}$ の値. T は true, F は false をそれぞれ表す.

a_C	$\Pr\{A_C = a_C\}$
T	0.5
F	0.5

a_S	$\Pr\{A_S = a_S A_C = T\}$	$\Pr\{A_S = a_S A_C = F\}$
T	0.1	0.5
F	0.9	0.5
a_R	$\Pr\{A_R = a_R A_C = T\}$	$\Pr\{A_R = a_R A_C = F\}$
T	0.8	0.2
F	0.2	0.8

a_W	$\Pr\{A_W = a_W A_S = T, A_R = T\}$	$\Pr\{A_W = a_W A_S = T, A_R = F\}$
T	0.99	0.9
F	0.01	0.1
a_W	$\Pr\{A_W = a_W A_S = F, A_R = T\}$	$\Pr\{A_W = a_W A_S = F, A_R = F\}$
T	0.9	0
F	0.1	1

式 (91) における $\Pr\{A_W = a_A | A_S = a_S, A_R = a_R\}$, $\Pr\{A_R = a_R | A_C = a_C\}$, $\Pr\{A_S = a_S | A_C = a_C\}$, $\Pr\{A_C = a_C\}$ の値は表 1 に与える. さて, ここで「芝生がぬれている」のは一体何が原因なのかを推定したいとする. この場合,

$$\begin{aligned}\Pr\{A_R = a_R | A_W = a_W\} &= \frac{\Pr\{A_R = a_R, A_W = a_W\}}{\Pr\{A_W = a_W\}} \\ \Pr\{A_S = a_S | A_W = a_W\} &= \frac{\Pr\{A_S = a_S, A_W = a_W\}}{\Pr\{A_W = a_W\}}\end{aligned}\quad (92)$$

を計算して比較することになる. これらの条件付き確率は次の周辺確率を計算することにより得られる.

$$\Pr\{A_R = a_R, A_W = a_W\} = \sum_{a_C=T, F} \sum_{a_S=T, F} \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \quad (93)$$

$$\Pr\{A_S = a_S, A_W = a_W\} = \sum_{a_C=T, F} \sum_{a_R=T, F} \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \quad (94)$$

$$\Pr\{A_W = a_W\} = \sum_{a_C=T, F} \sum_{a_S=T, F} \sum_{a_R=T, F} \Pr\{A_C = a_C, A_S = a_S, A_R = a_R, A_W = a_W\} \quad (95)$$

今の場合, $\Pr\{A_S = \text{true}, A_W = \text{ture}\} = 0.2781$, $\Pr\{A_R = \text{true}, A_W = \text{true}\} = 0.4581$, $\Pr\{A_W = \text{true}\} = 0.6471$ なので

$$\Pr\{A_R = \text{true} | A_W = \text{true}\} = \frac{\Pr\{a_R = \text{true}, A_W = \text{true}\}}{\Pr\{A_W = \text{true}\}} = \frac{0.4581}{0.6471} = 0.7079$$

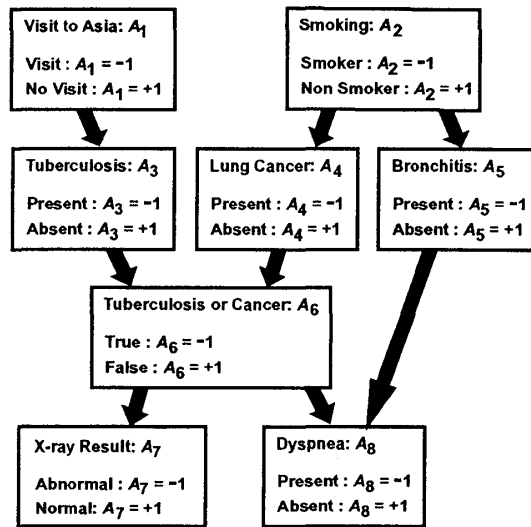


図 13: 確率的推論機構の例.

$$\Pr\{A_S = \text{true} | A_W = \text{true}\} = \frac{\Pr\{A_S = \text{true}, A_W = \text{true}\}}{\Pr\{A_W = \text{true}\}} = \frac{0.2781}{0.6471} = 0.4298 \quad (96)$$

これにより, 「芝生がぬれている」のは「雨が降った」ためだったということが推論できるわけである.

もう少し複雑な推論機構として図 13 に示す推論機構を考えよう. この例題はベイジアンネットの標準的な例題を掲載したホームページ <http://www.norsys.com/networklibrary.html> に「Asia」という名前で掲載されているものであり, もともとの出典は文献 [25] とされている. この推論機構は 8 個のノードから構成されている. 各ノードに 1 番から 8 番までの通し番号をふる. これにより図 13 の確率的推論機構は図 14 のようなグラフ表現として与えられる. 各ノードには前の例と同じように「+1」と「-1」という 2 つの状態が割り当てられている. この時, 結合確率分布はベイズ規則をもとに次のように構成される.

$$\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6, A_7 = a_7, A_8 = a_8\}$$

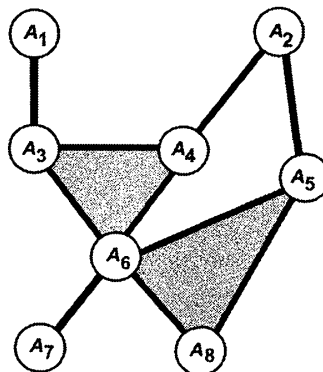


図 14: 図 13 の確率的推論機構のグラフ表現.

$$\begin{aligned}
 &= \Pr\{A_7 = a_7, A_8 = a_8 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\} \\
 &\quad \times \Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\} \\
 &= \Pr\{A_7 = a_7, A_8 = a_8 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\} \\
 &\quad \times \Pr\{A_5 = a_5, A_6 = a_6 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} \\
 &\quad \times \Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} \\
 &= \Pr\{A_7 = a_7, A_8 = a_8 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\} \\
 &\quad \times \Pr\{A_5 = a_5, A_6 = a_6 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} \\
 &\quad \times \Pr\{A_3 = a_3, A_4 = a_4 | A_1 = a_1, A_2 = a_2\} \\
 &\quad \times \Pr\{A_1 = a_1, A_2 = a_2\} \tag{97}
 \end{aligned}$$

最後の表式に見られる条件付き確率は図 14 では

$$\left\{ \begin{aligned}
 &\Pr\{A_7 = a_7, A_8 = a_8 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6\} \\
 &\quad = V_{56 \rightarrow 8}(a_8 | a_5, a_6) V_{6 \rightarrow 7}(a_7 | a_6) \\
 &\Pr\{A_5 = a_5, A_6 = a_6 | A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4\} \\
 &\quad = V_{2 \rightarrow 5}(a_5 | a_2) V_{34 \rightarrow 6}(a_6 | a_3, a_4) \\
 &\Pr\{A_3 = a_3, A_4 = a_4 | A_1 = a_1, A_2 = a_2\} = V_{1 \rightarrow 3}(a_3 | a_1) V_{2 \rightarrow 4}(a_4 | a_2) \\
 &\Pr\{A_1 = a_1, A_2 = a_2\} = V_1(a_1) V_2(a_2)
 \end{aligned} \right. \tag{98}$$

という形に与えられる. これにより式 (97) に与えられた結合確率分布は次のような形に書き換えられる.

$$\begin{aligned}
 &\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6, A_7 = a_7, A_8 = a_8\} \\
 &\quad = V_{56 \rightarrow 8}(a_8 | a_5, a_6) V_{6 \rightarrow 7}(a_7 | a_6) V_{34 \rightarrow 6}(a_6 | a_3, a_4) \\
 &\quad \quad \times V_{2 \rightarrow 5}(a_5 | a_2) V_{2 \rightarrow 4}(a_4 | a_2) V_{1 \rightarrow 3}(a_3 | a_1) V_2(a_2) V_1(a_1) \tag{99}
 \end{aligned}$$

ここで, 新たに

$$\left\{ \begin{aligned}
 &W_{568}(a_5, a_6, a_8) = V_{56 \rightarrow 8}(a_8 | a_5, a_6) \quad W_{67}(a_6, a_7) = V_{6 \rightarrow 7}(a_7 | a_6) \\
 &W_{346}(a_3, a_4, a_6) = V_{34 \rightarrow 6}(a_6 | a_3, a_4) \quad W_{25}(a_2, a_5) = V_{2 \rightarrow 5}(a_5 | a_2) V_2(a_2) \\
 &W_{24}(a_2, a_4) = V_{2 \rightarrow 4}(a_4 | a_2) V_2(a_2) \quad W_{13}(a_1, a_3) = V_{1 \rightarrow 3}(a_3 | a_1) V_1(a_1) \\
 &W_2(a_2) = V_2(a_2) \quad W_3(a_3) = W_4(a_4) = W_5(a_5) = W_6(a_6) = 1
 \end{aligned} \right. \tag{100}$$

という因子を導入し, 更に式 (99) は

$$\begin{aligned}
 &\Pr\{A_1 = a_1, A_2 = a_2, A_3 = a_3, A_4 = a_4, A_5 = a_5, A_6 = a_6, A_7 = a_7, A_8 = a_8\} \\
 &\quad = \frac{1}{Z} \left(\frac{W_{568}(a_5, a_6, a_8) W_{67}(a_6, a_7) W_{346}(a_3, a_4, a_6) W_{25}(a_2, a_5) W_{24}(a_2, a_4) W_{13}(a_1, a_3)}{W_2(a_2)^2 W_3(a_3) W_4(a_4) W_5(a_5) W_6(a_6)} \right) \tag{101}
 \end{aligned}$$

という形に書き直される. ここで, Z は規格化定数である. 更に新しい記号として

$$C \equiv \{568, 67, 346, 25, 24, 13, 2, 3, 4, 5, 6\} \tag{102}$$

$$\begin{cases} \mu(568) = \mu(67) = \mu(346) = \mu(25) = \mu(24) = \mu(13) = -1 \\ \mu(2) = 2, \quad \mu(3) = \mu(4) = \mu(5) = \mu(6) = 1 \end{cases}, \quad (103)$$

$$A_\alpha \equiv \{A_i | i \in \alpha\}, \quad a_\alpha \equiv \{a_i | i \in \alpha\} \quad (\alpha \in C) \quad (104)$$

を導入する. C の各要素はそれぞれがノードの集合であり,

$$\begin{aligned} 2 < 24, \quad 2 < 25, \quad 3 < 13, \quad 3 < 346, \\ 4 < 24, \quad 4 < 346, \quad 5 < 25, \quad 5 < 568, \\ 6 < 346, \quad 6 < 67, \quad 6 < 568, \end{aligned} \quad (105)$$

という包含関係にもとづく順序関係が存在し, C は半順序集合となっている. $i < \alpha$ はノード i が 2 個以上のノードからなる集合 α の要素の一つであることを示している. すなわち i は α の新部分集合となっているということもできる. ここで定義されたノードの集合 α はクラスターとも呼ばれる. これらの記号を用いると式 (99) は

$$\Pr\{\mathbf{A} = \mathbf{a}\} = \frac{\prod_{\alpha \in C} W_\alpha(a_\alpha)^{-\mu(\alpha)}}{\sum_{\mathbf{z}} \prod_{\alpha \in C} W_\alpha(z_\alpha)^{-\mu(\alpha)}} \quad (106)$$

$$\sum_{\mathbf{z}} \equiv \prod_{i=1}^8 \sum_{z_i = \pm 1} \quad (107)$$

というコンパクトな形にまとめられる.

$W_\alpha(a_\alpha)$ と $\mu(\alpha)$ から関数 $E(\alpha)$:

$$E(\mathbf{a}) \equiv - \sum_{\alpha \in C} \mu(\alpha) \ln(W_\alpha(a_\alpha)) \quad (108)$$

を導入すると, $\Pr\{\mathbf{A} = \mathbf{a}\}$ はギブス分布

$$\Pr\{\mathbf{A} = \mathbf{a}\} = \frac{\exp(-E(\mathbf{a}))}{\sum_{\mathbf{z}} \exp(-E(\mathbf{z}))} \quad (109)$$

として表現されることから, $E(\mathbf{a})$ はハミルトニアンとなっていることがわかる. 第 3 節の説明と同様にして $\Pr\{\mathbf{A} = \mathbf{a}\}$ は次のような形の自由エネルギー最小の変分原理を満たすことが確かめられる.

$$\Pr\{\mathbf{A} = \mathbf{a}\} = \min_Q \mathcal{F}[Q] \quad (110)$$

$$\mathcal{F}[Q] \equiv \mathcal{E}[Q] - \mathcal{S}[Q] \quad (111)$$

$$\mathcal{E}[Q] \equiv \sum_{\mathbf{z}} E(\mathbf{z}) Q(\mathbf{z}), \quad \mathcal{S}[Q] \equiv - \sum_{\mathbf{z}} Q(\mathbf{z}) \ln Q(\mathbf{z}) \quad (112)$$

ここで、試行関数としての確率分布 $Q(f)$ に対して周辺確率分布

$$Q_\alpha(a_\alpha) \equiv \sum_z \left(\prod_{\{i|i \in \alpha\}} \delta_{a_i, z_i} \right) Q(z) \quad (113)$$

を導入すると,

$$\sum_{z_\alpha} Q_\alpha(z_\alpha) = 1 \quad (\alpha \in C) \quad (114)$$

$$Q_\alpha(a_\alpha) = \sum_{z_\gamma} \left(\prod_{\{i|i \in \alpha\}} \delta_{a_i, z_i} \right) Q_\gamma(z_\gamma) \quad (\alpha < \gamma \in C) \quad (115)$$

$$\sum_{z_\alpha} \equiv \prod_{\{i|i \in \alpha\}} \sum_{z_i = \pm 1} \quad (116)$$

が成り立つことは容易に確かめられる. 更に, 式 (113) を用いると $\mathcal{E}[Q]$ は周辺確率分布 $Q_\alpha(a_\alpha)$ ($\alpha \in C$) だけを用いて

$$\mathcal{E}[Q] = \sum_{\alpha \in C} \mu(\alpha) \sum_{z_\alpha} Q_\alpha(z_\alpha) \ln(W_\alpha(z_\alpha)) \quad (117)$$

という形に表される.

クラスター変分法においては $Q(a)$ を

$$Q(a) = \frac{Q_{568}(a_5, a_6, a_8) Q_{67}(a_6, a_7) Q_{346}(a_3, a_4, a_6) Q_{25}(a_2, a_5) Q_{24}(a_2, a_4) Q_{13}(a_1, a_3)}{Q_2(a_2)^2 Q_3(a_3) Q_4(a_4) Q_5(a_5) Q_6(a_6)} \quad (118)$$

すなわち

$$Q(a) = \prod_{\alpha \in C} Q_\alpha(a_\alpha)^{-\mu(\alpha)} \quad (119)$$

という形に近似的に制限し, 式 (114) および式 (115) を拘束条件として自由エネルギー最小の変分原理 (110) のもとで決定された $Q_\alpha(a_\alpha) = \hat{Q}_\alpha(a_\alpha)$ が周辺確率 $\Pr\{A_\alpha = a_\alpha\}$ の近似として採用される. エントロピー $\mathcal{S}[Q]$ は式 (113) および式 (119) から次のように書き直される.

$$\mathcal{S}[Q] = \sum_{\alpha \in C} \mu(\alpha) \sum_{z_\alpha} Q_\alpha(z_\alpha) \ln(Q_\alpha(z_\alpha)) \quad (120)$$

従って自由エネルギー (111) は次のように書き下される.

$$\mathcal{F}[Q] \simeq \mathcal{F}_{\text{CVM}}[\{Q_\alpha | \alpha \in C\}] \equiv \sum_{\alpha \in C} \mu(\alpha) \sum_{z_\alpha} Q_\alpha(z_\alpha) \ln\left(\frac{Q_\alpha(z_\alpha)}{Q_\alpha(z_\alpha)}\right) \quad (121)$$

$\mathcal{F}[Q]$ を最小化する $\{Q_\alpha | \alpha \in C\}$ は近似自由エネルギー $\mathcal{F}_{\text{CVM}}[\{Q_\alpha | \alpha \in C\}]$ の条件付き変分原理として次のように与えられる.

$$\begin{aligned} \{\hat{Q}_\alpha | \alpha \in C\} &\simeq \arg \min_{\{Q_\alpha | \alpha \in C\}} \left\{ \mathcal{F}_{\text{CVM}}[\{Q_\alpha | \alpha \in C\}] \middle| \sum_{z_\alpha} Q_\alpha(z_\alpha) = 1 \quad (\alpha \in C), \right. \\ &\quad \left. Q_\alpha(a_\alpha) = \sum_{z_\gamma} \left(\prod_{\{i|i \in \alpha\}} \delta_{a_i, z_i} \right) Q_\gamma(z_\gamma) \quad (\alpha < \gamma \in C) \right\}. \end{aligned} \quad (122)$$

各拘束条件に対してラグランジュの未定係数を導入し、極値の条件を求め、その後、導入したラグランジュの未定係数を拘束条件を満たすように決めるという手順において、 $\{\hat{Q}_\alpha | \alpha \in C\}$ に対する決定方程式は次のように与えられる。

$$\hat{Q}_i(a_i) = \frac{W_i(a_i) \prod_{\{\alpha | \alpha \in c_i\}} \mathcal{M}_{\alpha \rightarrow i}(a_i)}{\sum_{z_i} W_i(z_i) \prod_{\{\alpha | \alpha \in c_i\}} \mathcal{M}_{\alpha \rightarrow i}(z_i)}, \quad (123)$$

$$\hat{Q}_\alpha(a_\alpha) = \frac{W_\alpha(a_\alpha) \prod_{\{i | i \in \alpha\}} \prod_{\{\gamma | \gamma \in c_i \setminus \alpha\}} \mathcal{M}_{\gamma \rightarrow i}(a_i)}{\sum_{z_\alpha} W_\alpha(z_\alpha) \prod_{\{i | i \in \alpha\}} \prod_{\{\gamma | \gamma \in c_i \setminus \alpha\}} \mathcal{M}_{\gamma \rightarrow i}(z_i)}, \quad (124)$$

$$\mathcal{M}_{\alpha \rightarrow i}(a_i) = \frac{\sum_{z_\alpha} \delta_{a_i, z_i} \left(\frac{W_\alpha(z_\alpha)}{W_i(z_i)} \right) \prod_{\{j | j < \alpha\}} \prod_{\{\gamma | \gamma \in c_j \setminus \alpha\}} \mathcal{M}_{\gamma \rightarrow j}(z_j)}{\sum_{z_\alpha} \left(\frac{W_\alpha(z_\alpha)}{W_i(z_i)} \right) \prod_{\{j | j < \alpha\}} \prod_{\{\gamma | \gamma \in c_j \setminus \alpha\}} \mathcal{M}_{\gamma \rightarrow j}(z_j)}. \quad (125)$$

このような取り扱いはどこまで一般化できるのか？ 気になる話である。ある種の条件の下で与えられた任意の推論システムでこの決定方程式が成り立てば、汎用性のある推論ソフトを作ることができる。実はこのからくりの根底には次の性質がある。

- (i) C はノードの集合を要素とする集合である。
- (ii) すべてのノードからなる集合を Ω とすると集合 C から集合 Ω を除いた集合 $C \setminus \Omega$ の任意の要素 γ と γ' に対して $\gamma < \gamma'$ または $\gamma > \gamma'$ という関係が成り立つことはない。
- (iii) $\mu(\gamma)$ は $\mu(\gamma) = -1$ ($\gamma \in C \setminus \Omega$), $\mu(i) = -1 - \sum_{\{\gamma | \gamma > i, \gamma \in C\}} \mu(\gamma)$ ($i \in \Omega$) により定義される。
- (iv) 全ノードに対する結合確率分布 $\Pr\{A = a\}$ が式 (106) の形に与えられる。

具体的に図 13 および図 14 で与えられた確率的推論について $V_{56 \rightarrow 8}(a_8 | a_5, a_6)$, $V_{34 \rightarrow 6}(a_6 | a_3, a_4)$, $V_{6 \rightarrow 7}(a_7 | a_6)$, $V_{2 \rightarrow 5}(a_5 | a_2)$, $V_{2 \rightarrow 4}(a_4 | a_2)$, $V_{1 \rightarrow 3}(a_3 | a_1)$, $V_1(a_1)$, $V_2(a_2)$ の値を表 2 のように設定し、式 (123)-(125) を解くことにより得られた周辺確率分布 ($\hat{Q}_i(+1)$, $\hat{Q}_i(-1)$) と定義 (113) にもとづいて得られた厳密な周辺確率分布 ($P_i(+1)$, $P_i(-1)$) の値を表 3 に与える。更に周辺確率分布 $\hat{Q}_{25}(a_2, a_5)$, $\hat{Q}_{36}(a_3, a_6)$ および条件付き確率 $\Pr\{A_2 = a_2 | A_5 = a_5\} \simeq \hat{Q}_{25}(a_2, a_5) / \hat{Q}_5(a_5)$, $\Pr\{A_3 = a_3 | A_6 = a_6\} \simeq \hat{Q}_{36}(a_3, a_6) / \hat{Q}_6(a_6)$ の値を表 4 および表 5 に与える。

ここで紹介したクラスター変分法を用いた確率的推論アルゴリズムの一部は実は人工知能の分野では確率伝搬法という名前で統計力学とはまったく独立に研究されていた。最近になってニューロコンピューティング分野の何人かの研究者によって、自由エネルギーの概念とクラスター変分法の導入が確率伝搬法のアルゴリズムの総括的理解につながるということが指摘されている [26, 27]。本節で紹介した例において、直接、因果関係で結ばれたノード間の推論は計算できることは説明したが、例えば 1

と 8 のような, 直接, 因果関係で結ばれていないようなノード間の推論の計算については触れなかった. これを可能にするためには, クラスター変分法に加えて, 線形応答理論というもう一つの新しい概念を導入する必要がある. 詳細は参考文献 [28] を参照していただきたい.

表 2: 式 (99) における $V_{56 \rightarrow 8}(a_8|a_5, a_6)$, $V_{34 \rightarrow 6}(a_6|a_3, a_4)$, $V_{6 \rightarrow 7}(a_7|a_6)$, $V_{2 \rightarrow 5}(a_5|a_2)$, $V_{2 \rightarrow 4}(a_4|a_2)$, $V_{1 \rightarrow 3}(a_3|a_1)$, $V_1(a_1)$, $V_2(a_2)$ の値.

a_1	V_1	a_2	V_2
+1	0.99	+1	0.50
-1	0.01	-1	0.50

a_1	a_3	$V_{1 \rightarrow 3}$	a_2	a_4	$V_{2 \rightarrow 4}$
+1	+1	0.99	+1	+1	0.99
+1	-1	0.01	+1	-1	0.01
-1	+1	0.95	-1	+1	0.90
-1	-1	0.05	-1	-1	0.10

a_2	a_5	$V_{2 \rightarrow 5}$	a_6	a_7	$V_{6 \rightarrow 7}$
+1	+1	0.70	+1	+1	0.95
+1	-1	0.30	+1	-1	0.05
-1	+1	0.40	-1	+1	0.02
-1	-1	0.60	-1	-1	0.98

a_3	a_4	a_6	$V_{34 \rightarrow 6}$	a_5	a_6	a_8	$V_{56 \rightarrow 8}$
+1	+1	+1	0.99	+1	+1	+1	0.90
+1	+1	-1	0.01	+1	+1	-1	0.10
+1	-1	+1	0.01	+1	-1	+1	0.30
+1	-1	-1	0.99	+1	-1	-1	0.70
-1	+1	+1	0.01	-1	+1	+1	0.20
-1	+1	-1	0.99	-1	+1	-1	0.80
-1	-1	+1	0.01	-1	-1	+1	0.10
-1	-1	-1	0.99	-1	-1	-1	0.90

表 3: 式 (123)-(125) を解くことにより得られた周辺確率分布 $(\hat{Q}_i(+1), \hat{Q}_i(-1))$ と定義 (113) にもとづいて得られた厳密な周辺確率分布 $(P_i(+1), P_i(-1))$ の値

i	$(\hat{Q}_i(+1), \hat{Q}_i(-1))$	$(P_i(+1), P_i(-1))$
1	(0.9900, 0.0100)	(0.9900, 0.0100)
2	(0.5000, 0.5000)	(0.5000, 0.5000)
3	(0.9896, 0.0104)	(0.9896, 0.0104)
4	(0.9450, 0.0550)	(0.9450, 0.0550)
5	(0.5500, 0.4500)	(0.5500, 0.4500)
6	(0.9265, 0.0735)	(0.9265, 0.0735)
7	(0.8816, 0.1184)	(0.8816, 0.1184)
8	(0.5574, 0.4426)	(0.5607, 0.4393)

表 4: 式 (123)-(125) を解くことにより得られた周辺確率分布 $\hat{Q}_{25}(a_2, a_5)$ および条件付き確率 $\Pr\{A_2 = a_2 | A_5 = a_5\} \simeq \hat{Q}_{25}(a_2, a_5) / \hat{Q}_5(a_5)$ の値

a_2	a_5	$\hat{Q}_{25}(a_2, a_5)$	$\Pr\{A_2 = a_2 A_5 = a_5\}$
+1	+1	0.35	0.70
+1	-1	0.15	0.30
-1	+1	0.20	0.40
-1	-1	0.30	0.60

表 5: 式 (123)-(125) を解くことにより得られた周辺確率分布 $\hat{Q}_{36}(a_3, a_6)$ および条件付き確率 $\Pr\{A_3 = a_3 | A_6 = a_6\} \simeq \hat{Q}_{36}(a_3, a_6) / \hat{Q}_6(a_6)$ の値

a_3	a_6	$\hat{Q}_{36}(a_3, a_6)$	$\Pr\{A_3 = a_3 A_6 = a_6\}$
+1	+1	0.9264	0.9361
+1	-1	0.0632	0.0639
-1	+1	0.0001	0.0120
-1	-1	0.0103	0.9880

9 おわりに

本講義では、自由エネルギーの変分原理にもとづく確率的情報処理アルゴリズムについての概説を行った。本講義が対象としている研究分野はまだ始まったばかりの若い分野であり、研究者の絶対数もそれほど多くない。したがって、誰も手をつけたことのない研究テーマが山積している。今、参入すれば「この問題ならあの人に聞くのが一番と言われる人」あるいは「日本での、あるいは世界での第一人者」になるチャンスかもしれない。この機会に「統計力学との関連にもとづく確率的情報処理の研究に参入してみよう」と1人でも多くの若手研究者の皆さんにお考えいただけたら幸いである。統計力学を情報処理に応用する最新の研究成果の一部は科研費特定領域研究「確率的情報処理への統計力学的アプローチ」(平成14年度から平成17年度までの4年間)のホームページ (URL: <http://www.smapip.eei.metro-u.ac.jp/>) から得られるので是非参照していただきたい。

謝辞 本講義の準備の過程で様々な面で助言をいただいた東北大学大学院理学研究科物理学専攻の藤田早苗さんに深く感謝いたします。

参考文献

- [1] 樺島祥介, “コトの物理 — 誤り訂正符号を例として —,” 日本物理学会誌, vol.58, no.4, pp.239-246, 2003.
- [2] Y. Kabashima, T. Murayama and D. Saad, “Cryptographical Properties of Tsing Spin Systems,” Physical Review Letters, vol.84, no.9, pp.2030-2033, 2000.
- [3] T. Murayama, “Statistical mechanics of data compression theorem,” Journal of Physics A: Mathematical and General, vol.35, no.8, pp.L95-L100, 2002.
- [4] R. Chellappa and A. Jain (eds), *Markov Random Fields: Theory and Applications*, Academic Press, New York, 1993.
- [5] S. Z. Li, *Markov Random Field Modeling in Computer Vision*, Springer-Verlag, Tokyo, 1995.
- [6] K. Tanaka, “Statistical-mechanical approach to image processing,” Journal of Physics A: Mathematical and General, vol.35, no.20, pp.R81-R150, 2002.
- [7] 田中和之, “統計力学的手法をもとにした画像修復,” 日本物理学会誌, vol.54, no.1, pp.25-33, 1999.
- [8] 田中和之, “量子力学的に拡張されたライン場をもつ結合ガウス・マルコフ確率場モデルを用いた画像修復,” 電子情報通信学会論文誌 (D-II), vol.J84-D-II, no.4, pp.737-743, 2001
- [9] 本村陽一, “ベイジアンネットソフトウェア,” 人工知能学会誌, vol.17, no.5, pp.559-565, 2002.
- [10] 池田思朗, 田中利幸, M. Welling, NIPS*2002 post conference workshop on Propagation Algorithms on Graphs with Cycles: Theory and Applications (December 13th and 14th, 2002, Whistler, Canada), URL: <http://prop.algo.eei.metro-u.ac.jp/>.

- [11] 田中利幸, “移動体通信技術とスピングラスとの意外な関係,” 日本物理学会誌, vol.56, no.9, pp. 660-666, 2001.
- [12] T. Tanaka, “A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors,” IEEE Transactions on Information Theory, vol.48, no.11, pp.2888- 2910, 2002.
- [13] 村山立人, 樺島祥介, “スピングラス理論と機械学習,” 人工知能学会誌, vol.16, no.5, pp.683-688, 2001
- [14] 村田昇編, “特集/情報論的学習理論 —機械学習のさまざまな形—,” Computer Today (2003 年 3 月号), 2003.
- [15] 三宮信夫, 喜多一, 玉置久, 岩本貴史, “遺伝的アルゴリズムと最適化,” 朝倉書店, 1998.
- [16] T. Horiguchi, “Routing control of packet flow using neural network,” Physica A, vol.297, pp.521-531, 2001.
- [17] O. Watanabe, “Simple sampling technique for discovery science,” IEICE Transactions on Information and Systems, vol.E83-D, no.1, pp.19-26, 2000.
- [18] M. Oppen and D. Saad (eds), “Advanced Mean Field Methods — Theory and Practice —,” MIT Press, 2001.
- [19] 樺島祥介, “岩波講座 物理の世界/学習と情報の平均場理論,” 岩波書店, 2002.
- [20] 特集/知識情報処理の統計力学的アプローチ, 数理科学, 1999 年 12 月号.
- [21] 西森秀稔, “スピングラス理論と情報統計物理学,” 新物理学選書, 岩波書店, 1999.
- [22] H. Nishimori, *Statistical Physics of Spin Glasses and Information Processing: An Introduction*, Oxford University Press, Oxford, 2001.
- [23] J. Pearl: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*, (Morgan Kaufmann, 1988).
- [24] J. Pearl, *Causality*, Cambridge. 2000.
- [25] S. L. Lauritzen and D. J. Spiegelhalter, “Local computations with probabilities on graphical structures and their application to expert systems,” J. Royal Statistics Society B, vol.50, no.2, pp.157-194, 1988.
- [26] J. S. Yedidia, W. T. Freeman and Y. Weiss: “Generalized belief propagation,” Advances in Neural Information Processing Systems, vol.13, pp.689-695, 2001 (Cambridge, MA: MIT Press).
- [27] H. J. Kappen and W. Wiegerinck: “Novel iteration schemes for the cluster variation method,” Advances in Neural Information Processing System, vol.14, pp.415-422, 2002 (Cambridge, MA: MIT Press).
- [28] K. Tanaka, “Probabilistic Inference by means of Cluster Variation Method and Linear Response Theory,” IEICE Transactions on Information and Systems, vol.E86-D, no.7, pp.1228-1242, 2003.